

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ЧЕРКЕСОВА Полина Андреевна

Выпускная квалификационная работа

**Статистический анализ больших данных для оценки динамических
характеристик в информационных системах оценочных компаний**

Направление 38.04.05 «Бизнес-информатика»
Основная образовательная программа магистратуры
ВМ.5604 «Информационная бизнес-аналитика»

Научный руководитель: СПбГУ,
к.ф.-м.н., доцент ЛАСКИН
Михаил Борисович

Рецензент: заведующий
Лабораторией ИТ на транспорте,
СПИИРАН, д.т.н., профессор
ИСКАНДЕРОВ Юрий Марсович

Санкт-Петербург
2020

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ ЦЕН НА РЫНКЕ НЕДВИЖИМОСТИ.....	6
1.1. Рыночная и кадастровая стоимость. Оценка и прогноз.....	6
1.2. Методы оценки и прогнозирования рыночной стоимости в российском и зарубежном научном сообществе	10
ГЛАВА 2. АНАЛИЗ НА ОСНОВЕ СРАВНЕНИЯ РЫНОЧНОЙ И КАДАСТРОВОЙ СТОИМОСТИ ОБЪЕКТОВ.....	22
2.1. Наиболее вероятная траектория рыночной стоимости во времени.....	22
2.2. Анализ временных рядов.....	42
ГЛАВА 3. ПРОГНОЗИРОВАНИЕ РЫНОЧНОЙ СТОИМОСТИ.....	47
3.1. Прогнозы на основе моделей временных рядов.....	47
ЗАКЛЮЧЕНИЕ	62
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	65

ВВЕДЕНИЕ

В рыночной экономике недвижимость занимает особое место, поскольку одновременно выполняет ряд уникальных функций и является частью нескольких важных сфер жизнедеятельности и бизнеса. Коммерческая недвижимость входит в состав средств производства, является базой для хозяйственной деятельности и развития для предприятий, жилая недвижимость выступает основой личного имущества граждан и может использоваться ими для непроектованных нужд.

Актуальность исследования данной области обусловлена тем, что развитие рынка недвижимости напрямую зависит от принятия решений, основанных на его исследовании и анализе. Они востребованы его участниками и, в том числе, государством. Результаты анализа важны для принятия управленческих решений со стороны отдельных потребителей и заинтересованных структур, реализации эффективной государственной политики в области инфраструктуры и различных жилищных проектов, корректировки и оценки показателей при ипотечном кредитовании. Таким образом, в прогнозировании рыночной стоимости объектов недвижимости заинтересованы владельцы различных форм собственности, строительные и девелоперские компании, страховые компании, банки, профессиональные участники рынка недвижимости, такие как организаторы торгов, риэлторы.

Объектом исследования выступают динамические характеристики объектов жилой недвижимости.

Предметом исследования выступают прогнозные значения рыночных цен при определённой кадастровой стоимости. В качестве базы для сравнительного анализа рыночных изменений выбрана кадастровая стоимость, так как наиболее полные данные по рынку содержатся в кадастровых базах, в которые, в том числе, занесена оценка рыночной стоимости на определённую дату.

Целью данного исследования является получение прогноза на основе систематических сравнений распределений кадастровых стоимостей и рыночных данных о ценах внутри достаточно однородных кластеров объектов недвижимости и построения наиболее вероятной траектории движения рыночной цены для любых объектов недвижимости, даже для тех, которые за период наблюдений ни разу не попадали в листинги продающихся объектов.

Исходя из поставленной цели можно выделить следующие задачи:

1. Исследовать понятия рыночной и кадастровой стоимости, стандарты их оценки и специфику относительно российского рынка жилой недвижимости
2. Синтезировать основную информацию в научной литературе относительно методов оценки и прогнозирования рыночной цены на недвижимость
3. Осуществить сбор данных, их очистку и первичный анализ

4. Проанализировать совокупность данных, программное обеспечение для работы с ними и подходящие методы и приёмы для обработки и анализа.
5. Рассмотреть совместные распределения кадастровой и рыночной стоимостей, определить их статистическую зависимость и наиболее вероятную цену для значения кадастровой стоимости объекта. Построить ковариационную матрицу для многомерного совместно нормального распределения логарифмов кадастровой стоимости и цен на различные моменты времени.
6. Определить модель и методы для прогнозирования временных рядов и оценить их точность.
7. Проинтерпретировать результаты, сделать выводы, важные для оценочного бизнеса.

До недавнего времени область исследования являлась мало проработанной в силу проблем организационного и нормативного характера при оценке объектов недвижимости, а также особенностей данных, таких как неструктурированность, большой объём, которые делают процесс обработки и анализа сложным и требующим определённого программного обеспечения.

С 2015 года при введении в эксплуатацию кадастровой стоимости в Российской Федерации появилась возможность исследования рынка на основе кадастровых баз. Кроме того, в последнее время в большинстве своём в зарубежных исследованиях стали применяться прогрессивные методы оценки и прогнозирования стоимости объектов недвижимости. Это связано с появлением соответствующего программного обеспечения.

Данное исследование базируется на сопоставлении рыночной и кадастровой стоимости и дальнейшем анализе как многомерной случайной величины. При изучении материалов и публикаций данной области не было встречено работ, в которых бы сравнивались рыночная и кадастровая стоимость.

Теоретической значимостью работы является алгоритм оценки наиболее вероятной траектории рыночной стоимости исследуемого объекта на основе многомерного условного логарифмически нормального распределения цен при заданном значении кадастровой стоимости. Также метод имеет практический характер и может быть применен в автоматических системах поддержки принятия решений оценочных компаний.

Для удобства и логической обоснованности работа разделена на три части. В первой части раскрывается теоретическая сущность понятия рыночной и кадастровой стоимости. В ней проведён анализ научной литературы, рассмотрены основные методы оценки и прогнозирования рыночной стоимости в российском научном сообществе и в зарубежных исследованиях.

Во второй главе рассмотрены теоретические основы анализа временных рядов, их свойства и модели прогнозирования. Кроме того, описываются данные, использованные для исследования, процесс их обработки и анализа с учётом их размера и специфики. Также приведена общая формула, на которой базируется дальнейший анализ, общее утверждение, и определяется наиболее вероятная траектория рыночной стоимости на примере двух различных по своей структуре районов для двух моментов времени.

Третья глава имеет прикладной характер. В ней строятся прогнозы для двух районов города, различных по структуре жилья, на основе выбранной модели прогнозирования. Модели выбираются исходя из описанных в предыдущей главе критериев и методик. Полученные результаты интерпретируются с учётом текущих и ожидаемых настроений на рынке недвижимости.

ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ ЦЕН НА РЫНКЕ НЕДВИЖИМОСТИ

1.1. Рыночная и кадастровая стоимость. Оценка и прогноз

В 2015 году было принято решение проводить кадастровую оценку недвижимости с помощью государственных бюджетных учреждений по единой установленной методологии. С введением в Российской Федерации такого понятия как кадастровая стоимость появилась возможность анализировать ее как срез рынка. Каждому объекту недвижимости был присвоен кадастровый номер и кадастровая стоимость. Это формирует основу для создания баз данных, содержащих такие характеристики, как кадастровый номер, кадастровая стоимость, цена предложения, различные ценообразующие факторы.

Введение кадастровой стоимости в РФ открывает новые широкие возможности для анализа и прогнозирования рыночных цен, так как кадастровые базы содержат наиболее полные списки объектов недвижимости, включая кадастровую стоимость, которая теперь в соответствии с законодательством должна обновляться не реже, чем раз в три года и, по состоянию на 2015 и 2018 год, была определена как рыночная стоимость и, следовательно, до следующей кадастровой оценки может служить базой для постоянного сравнения с рыночными данными, которые все время меняются, прежде всего по составу объектов.

Эти понятия тесно связаны с деятельностью по оценке объектов недвижимости. Прогнозирование динамических характеристик также основывается на предварительной оценке объектов и является завершающим этапом в исследовании рынка.

Задачей оценки является определение одной из стоимости объекта до совершения сделки. Она бывает рыночной, инвестиционной, ликвидационной, кадастровой и др.

Приказом Министерства экономического развития от 20 июля 2007 года был установлен Федеральный стандарт оценки «Цель оценки и виды стоимости», сокращенное название – ФСО №2.

В нём раскрывается цель оценки объектов недвижимости – определить стоимость объекта в зависимости от предполагаемого использования результата. Также приведены понятия разных видов стоимостей. В данном исследовании задействованы рыночная и кадастровая стоимость.

В стандарте указано, что при определении кадастровой стоимости устанавливается рыночная стоимость, утверждённая методами массовой оценки, в основном для целей налогообложения.

Определение рыночной стоимости раскрывается в Федеральном Законе №135 от 29.07.1998 «Об оценочной деятельности в Российской Федерации». Под рыночной

стоимостью понимается «наиболее вероятная цена, по которой данный объект оценки может быть отчужден на открытом рынке в условиях конкуренции, когда стороны сделки действуют разумно, располагая всей необходимой информацией, а на величине цены сделки не отражаются какие-либо чрезвычайные обстоятельства».

Таким образом, совершенно естественным является предположение, что наблюдения по рыночной стоимости являются случайными величинами и необходимо применять к ним статистические методы анализа. Законы распределения заранее не известны, возникает проблема их подбора, оценки параметров и выбора числовых характеристик, на основании которых делаются выводы о величине рыночной стоимости. Из текста Федерального Закона №135 следует, что такой числовой характеристикой является наиболее вероятная точка, т.е. точка максимума плотности распределения.

Существует три классических подхода к оценке недвижимости:

- Сравнительный
- Затратный
- Доходный

Согласно Федеральному стандарту оценки "Общие понятия оценки, подходы к оценке и требования к проведению оценки" (далее ФСО №1) от 20 июля 2007 года, утверждённому приказом Министерства экономического развития: «Сравнительный подход - совокупность методов оценки стоимости объекта оценки, основанных на сравнении объекта оценки с объектами - аналогами объекта оценки, в отношении которых имеется информация о ценах. Объектом - аналогом объекта оценки для целей оценки признается объект, сходный объекту оценки по основным экономическим, материальным, техническим и другим характеристикам, определяющим его стоимость.»

Сравнительный подход основывается на экономическом принципе взаимодействия спроса и предложения. Суть подхода состоит в том, что он производит оценку рыночной стоимости на основе анализа рыночных цен сделок или предложений по продаже или аренде объектов, которые являются аналогами оцениваемого объекта и которые были завершены на рынке оцениваемого объекта в прошлом до даты оценки.

Сравнительный подход может быть использован в том случае, если на дату оценки объекта имеется информация о достаточном для оценки количестве сделок или предложений. В результате этого с помощью сравнительного подхода можно получить достаточно хорошие и достоверные результаты для оценки квартир в жилых домах или встроенных помещений коммерческого назначения. Метод практически никогда не применяют для объектов специального назначения, у которых нет рынка или их рынок является ограниченным, таких как церкви, школы, стадионы, поскольку в строительстве

задействовано государство и их рынок не строится на принципе рыночного спроса и предложения.

Сравнительный подход при наличии достаточного количества данных позволяет получить достоверные результаты, которые можно легко объяснить. Когда данных недостаточно для получения достоверных точечных оценок, сравнительный метод применяют для получения диапазона стоимостей, в котором с большой долей вероятности находится искомая случайная величина – рыночная стоимость.

При использовании сравнительного подхода в оценке часто возникает проблема выбора объектов сравнения. Элементы для сравнения берутся из выборок, сформированных по тому или иному принципу однородности. У многих оценщиков при выборе объектов для сравнения возникает вопрос о возможности выбора объектов в разных районах города, разных типов домов и другие. К тому же, на рынке недвижимости объекты сравнения далеко не всегда однородны и возникает необходимость в разделении смесей. Для ее решения в работе может быть использован кластерный анализ на основе ценообразующих факторов. При определении рыночной стоимости объектов недвижимости методом сравнительного подхода также возникают следующие проблемы: формирование достаточно представительной выборки, построение оценки рыночной стоимости на основе статистических данных.

В ФСО №1 даётся следующее определение затратного подхода: «Затратный подход — совокупность методов оценки стоимости объекта оценки, основанных на определении затрат, необходимых для воспроизводства либо замещения объекта оценки с учетом износа и устаревания.

Затратами на воспроизводство объекта оценки являются затраты, необходимые для создания точной копии объекта оценки с использованием применявшихся при создании объекта оценки материалов и технологий.

Затратами на замещение объекта оценки являются затраты, необходимые для создания аналогичного объекта с использованием материалов и технологий, применяющихся на дату оценки»

Метод базируется на одной из теорий ценообразования — теории трудовой стоимости, согласно которой стоимость товара определяется количеством и качеством труда и материальных ресурсов, потраченных на его создание. Затратный подход демонстрирует реальные затраты, необходимых для создания объекта недвижимости, —это сумма, необходимая для закупки необходимых материалов и найма рабочей силы для строительства объекта недвижимости. Недостатком метода является его неспособность объяснить проблему разной стоимости объектов недвижимости, имеющих одинаковый состав

материалов. Затратный подход не может учесть в полной мере соотношение сил спроса и предложения на аналогичные объекты недвижимости.

Согласно ФСО №1: «Доходный подход - совокупность методов оценки стоимости объекта оценки, основанных на определении ожидаемых доходов от использования объекта оценки.» Для использования доходного подхода к оценке объектов недвижимости в значительной степени необходима информация о ставке аренды, так как рынок аренды более динамично развивается по сравнению с рынком продаж жилья, и в его рамках совершается большее количество сделок.

Оценка рыночной стоимости недвижимости с использованием доходного подхода основана на преобразовании доходов, которые, как ожидается, имущество будет генерировать в процессе оставшейся экономической жизни в стоимость. Источник дохода должен быть продуктом оцениваемого актива, это может быть продажа, аренда, дивиденды, прибыль и др.

В рамках этого подхода различают метод прямой капитализации доходов и метод капитализации доходов по норме отдачи на капитал (метод капитализации потенциальных доходов)

Использование одного из этих подходов или методов оценки стоимости зависит от исходной рыночной информации, типа и категории объектов недвижимости.

Это классические подходы к оценке недвижимости. К жилой недвижимости по большей степени применяют сравнительный или доходный подход к оценке. Однако есть ряд работ в российском научном сообществе, которые используют для оценки показателей статистические методы на основе многомерного логарифмически нормального закона распределения.

Аналогично исследованиям финансового рынка для рынка недвижимости могут применяться два самостоятельных направления анализа – фундаментальный и технический.

Фундаментальный анализ подразумевает исследование рынка недвижимости на основе изучения влияния факторов, определяющих закономерности и тенденции его развития, то есть прогнозные цены определяются на основе изменений одного или нескольких факторов. Факторы могут быть:

- экономического— изменение макроэкономических показателей, арендных ставок, коммунальных тарифов, ставок по кредитам
- административного— земельный налог, налог на имущество, изменения в законодательстве
- социального — демографические показатели, уровень образования и преступности

- экологического характера – условия окружающей среды, развитие инфраструктуры и др.

Цель технического анализа – выработать подход к изучению рынка, прогнозы которого осуществляются на основе анализа наблюдаемых или прошлых тенденций его развития.

1.2. Методы оценки и прогнозирования рыночной стоимости в российском и зарубежном научном сообществе

Данные по объектам недвижимости представляют интерес для научного исследования в сформировавшихся в современном мире условиях цифровой экономики. Рыночные данные, как правило, являются плохо структурированными, могут быть отнесены к большим данным, в том смысле, что их объем, не позволяет без предварительной подготовки применять традиционные статистические методы. Наметилась тенденция внедрения в практику оценки и анализа рынков недвижимости инновационных, нестандартных подходов.

Российское оценочное сообщество в основном придерживается традиционных подходов к оценке стоимости недвижимости. Теория и практика оценки в РФ является относительно молодым направлением экономической мысли с момента принятия ФЗ №135 «Об оценочной деятельности в РФ» в 1998 году. Российские ученые внесли большой вклад в теорию оценки недвижимого имущества, среди наиболее значимых работ следует выделить труды Озерова Е.С., Грибовского С.В., Коростелева С.П.

В учебном пособии «Оценка стоимости недвижимости» Грибовский С.В. начинает с вопросов общей теории экономики недвижимости.¹ Методы оценки стоимости недвижимости рассматриваются в рамках трех известных классических подходов к оценке стоимости активов: затратного, сравнительного и доходного. В доходном подходе рассматриваются следующие методы оценки: прямая капитализация, дисконтирование денежных потоков, капитализация по расчётным моделям, капитализация с учётом заёмных средств. Каждый из подходов сопровождается соответствующим математическим и алгоритмическим обеспечением, позволяющим использовать их в практической деятельности по оценке объектов недвижимости. Методы излагаются для оценки стоимости всех типов объектов недвижимости: свободных и застроенных земельных участков, встроенных объектов недвижимости и объектов недвижимости, требующих определенных преобразований. Свою работу автор заканчивает изложением вопроса структуры отчета об объекте оценки недвижимости и требований к рецензированию. Методы из работы Грибовского С.В. следует применять на начальном этапе изучения ценообразования и

¹ Грибовский С.В. Оценка стоимости недвижимости: Учебное пособие. — 2-е изд., испр. и доп. — М.: ООО «Про-Аппрайзер» Онлайн, 2016. — 464 с.

оценки недвижимости для формирования представления о базе, на которой строится оценка объектов недвижимости Российского оценочного сообщества.

В работе «Математические методы оценки стоимости недвижимого имущества» Грибовский С.В и М.А. Федотова рассматривают теоретические основы и примеры применения различных математических методов при оценке имущества.² Авторы подробно изложили методологию и практику построения, анализа и использования математических моделей на базе корреляционно-регрессионного анализа, анализа временных рядов с использованием методов матричной алгебры. Также рассмотрена возможность применения математических методов и моделей при массовой оценке недвижимости.

В монографии Е.С.Озерова и С.В.Пупенцовой «Управление стоимостью и инвестиционным потенциалом недвижимости» представлен подход к анализу доходности недвижимости и к определению методологии управления изменениями с точки зрения обеспечения увеличения стоимости и инвестиционного потенциала (инвестиционной привлекательности) доходной недвижимости.³ В разделе экономического анализа обосновываются критерии отнесения имущества к недвижимости как экономическому благу и выделены особенности формирования критериев эффективности хозяйствования на объекте доходной недвижимости. Также отмечено, что в процедурах экономического анализа, целесообразно использовать инновации в методологии оценки. Выполнен анализ ценообразующих характеристик объектов недвижимости.

В то же время, представляются недостаточно проработанными методы сравнительного подхода в оценке недвижимого имущества, прежде всего потому, что возник разрыв между практикой оценки, основанной на малых выборках и наличием и доступностью больших рыночных данных. Анализ больших данных требует применения современных прикладных статистических пакетов.

В условиях наличия большого объема рыночных данных и данных кадастрового учета анализ изменений рыночной стоимости жилой недвижимости требует применения статистических методов исследования больших данных.

В недавно опубликованной статье Ласкина М. Б и Гадасиной Л.В. «Как определить кадастровую стоимость» указывается на кадастровую стоимость как важный инструмент мониторинга рынка и на взаимосвязь кадастровой стоимости (как рыночной, полученной на

² Грибовский С.В., Математические методы оценки стоимости недвижимого имущества: учеб. пособие / С.В. Грибовский, С.А. Сивец; под ред. С.В. Грибовского, М.А. Федотовой - М.: Финансы и статистика, 2008. - 368с.

³ Е.С. Озеров, С.В. Пупенцова., Управление стоимостью и инвестиционным потенциалом недвижимости - СПб : Изд-во Политехн. ун-та, 2015 . – 600 с.

конкретную дату методами массовой оценки) с текущей рыночной стоимостью.⁴ Подход опирается на стохастическую модель ценообразования, которая предполагает совместное нормальное распределение логарифмов цен. В данном случае этот подход применен к двумерному совместному распределению кадастровых стоимостей 2015 и 2018 года. Основная логика описывается следующим рассуждением:

Рассмотрим двумерную случайную величину $(V_{kc}, V_{цп})$ (V_{kc} – кадастровая стоимость, $V_{цп}$ – цена предложения). Пусть $V_{kc}, V_{цп}$ имеют совместное логарифмически нормальное распределение (т.е. их логарифмы распределены совместно нормально), с параметрами μ_1, σ_1 (для кадастровой стоимости) и μ_2, σ_2 (для цен предложений) соответственно. Тогда наиболее вероятная цена V_{pc} (рыночная стоимость) равна:

$$V_{pc} = Mode(V_{цп} | V_{kc} = v) = \exp(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (\ln(v) - \mu_2) - \sigma_1^2 (1 - \rho^2))$$

где ρ – коэффициент корреляции.

Эту формулу можно использовать для оценки текущей рыночной стоимости объекта (в приведенных единицах). Такая возможность появилась совсем недавно после завершения кадастрового учета и присвоения кадастровых стоимостей. Более того, это приводит к совершенно новой реальности в оценке – при создании обширных баз данных появится возможность по кадастровому номеру объекта немедленно, основываясь на рыночных данных, оценить его рыночную стоимость. Учет ценообразующих факторов в традиционном понимании трансформируется в процедуру формирования выборки (приблизительно однородных объектов), например при проведении кластеризации массива рыночных данных.

Таким образом зависимость рыночной стоимости (РС) от кадастровой имеет вид степенной функции:

$$V_{pc} = Mode(V_{цп} | V_{kc} = v) = Av^{\rho \frac{\sigma_1}{\sigma_2}}, \text{ где } A = \exp(\mu_1 - \rho \frac{\sigma_1}{\sigma_2} \mu_2 - \sigma_1^2 (1 - \rho^2))$$

Эта работа строится на методе оценки на основании стохастической модели ценообразования и соответствии эмпирических распределений цен логнормальному закону распределения вероятностей. Причины формирования логнормальной генеральной совокупности обоснованы в статье «Стохастическая модель ценообразования на рынке

⁴ Ласкин М.Б., Гадасина Л.В. Как определить кадастровую стоимость // Имущественные отношения в Российской Федерации. 2018. № 3. 42-53 с.

недвижимости: формирование логнормальной генеральной совокупности»⁵ Русаковым О.В., Ласкиным М.Б. и Джуксумбаевой О.И. и в статье «Особенности формирования величины рыночной стоимости недвижимости при логарифмически нормальном распределении цен»⁶ Ласкиным М.Б., Русаковым О.В., Джуксумбаевой О.И. и Ивакиной И.И. Такой подход может быть использован для оценки и прогнозирования рыночной стоимости на основе кадастровых данных и данных по рынку за разные периоды времени.

Для этого подхода необходимо проверить статистическую гипотезу о совместной нормальности логарифмов цен. Однако, традиционные тесты на нормальность не могут дать результата на выборках объёмом более 1000, т.к. существенно зависят от объема выборки. В работе Б.Ю. Лемешко, С.Б. Лемешко, М.А. Семеновой «К вопросу статистического анализа больших данных»⁷ предложен подход к решению проблемы больших выборок. Авторы отмечают, что при работе с большими данными целесообразно использовать методы оценивания параметров, предусматривающие группирование данных (они робастны, а вычислительные затраты не зависят от объёмов выборок), а непараметрические критерии согласия применять к выборкам, извлекаемым из больших данных, объём которых ограничивается точностью представления этих данных (количеством возможных уникальных значений в выборке).

Такие подходы к оценке рыночной стоимости существуют на данный момент в российском оценочном и научном сообществе.

Говоря о теоретических основах оценки недвижимости, нужно упомянуть работу британских учёных 2006 года.⁸ Авторы предлагают метод оценки имущества посредством изучения механизма ценообразования в сфере недвижимости с точки зрения бизнеса. В книге также анализируются ставки дисконтирования, аренда, покупательная способность, финансовые структуры, поведение инвесторов и все, что может окружать рынок недвижимости как экономическую структуру.

Американский профессор Асват Дамодаран посвятил много исследований оценке. В

⁵ Русаков О. В. Стохастическая модель ценообразования на рынке недвижимости: формирование логнормальной генеральной совокупности / О. В. Русаков, М. Б. Ласкин, О. И. Джаксумбаева // Вестник УМО. 2015. № 5.

⁶ Русаков О.В., Ласкин М.Б., Джаксумбаева О.И., Ивакина А.А. Особенности рыночной стоимости на рынке недвижимости при логарифмически нормальном распределении // Имущественные отношения в Российской Федерации. 2016. № 2(173). С. 40–50.

⁷ Лемешко Б. Ю. К вопросу статистического анализа больших данных / Б. Ю. Лемешко, С. Б. Лемешко, М. А. Семёнова // Вестник Томского государственного университета. Управление, вычислительная техника и информатика - 2018. – № 44. – С. 40–49.

⁸ Sayce S., Smith J., Cooper R., Venmore-Rowland P. (2009). Real Estate Appraisal: From Value to Worth. John Wiley & Sons. – 352 p.

одной из своих книг⁹ автор закладывает основы инвестиционной оценки в разных сферах экономики и бизнеса, в том числе и на рынке недвижимости. В работе рассмотрен подход дисконтированных денежных потоков, сравнительная и относительная оценка, ряд моделей и их особенности.

В зарубежных исследованиях предлагаются методы, которые отличаются от стандартных. Одним из них является метод гедонистического ценообразования. Суть метода заключается в том, что он оценивает влияние характеристик товара, которые косвенно влияют на его цену. Товар представляется как сумма отдельных атрибутов. Метод гедонистического ценообразования помогает оценить блага, которые не имеют стандартной рыночной оценки. Здесь важную роль играют предпочтения потребителей, эстетическая составляющая, дизайн, окружающая среда. Метод используется в исследовании потребителей и рынков, расчете индексов потребительских цен, оценке налогов, оценке автомобилей компьютеров. В последнее время подход активно используется для оценки рынка недвижимости.¹⁰

Метод заключается в обнаружении статистической зависимости между стоимостью жилой недвижимости и ценообразующими факторами, такими как общая площадь квартиры, жилая площадь, тип дома, расстояние до метро, расстояние до делового центра, количество комнат, этажность, качество окружающей среды и т.д.

В основном для оценки статистической зависимости применяются линейные, логарифмические или частично-логарифмические модели, но в последнее время идет активный поиск новых нетрадиционных подходов к анализу информации на рынке недвижимости. Это связано с появлением нового программного обеспечения, позволяющего обрабатывать значительные объёмы структурированных и неструктурированных «больших» данных и извлекать из них экономический смысл.

Ниже приведены наиболее интересные работы, описывающие методы оценки и прогнозирования рыночных цен на недвижимость.

Регрессионные модели оценки недвижимости.

Метод гедонистического ценообразования предлагает соответствующий подход для оценки таких внешних выгод, как благоустройство, обеспечиваемое городскими зелеными насаждениями, водоемами и хорошим качеством окружающей среды, которые имеют

⁹ Damodaran A. (2012). Investment Valuation: Tools and Techniques for Determining the Value of Any Asset (Vol. 666). John Wiley & Sons. – 992 p.

¹⁰ Herath, S. K. & Maier, G. (2010). The hedonic price method in real estate and housing market research. A review of the literature. Institute for Regional Development and Environment (pp. 1-21). Vienna, Austria: University of Economics and Business

существенное влияние на цены на недвижимость. В исследовании ¹¹ изучалось влияние ключевых элементов окружающей среды, влияющих на стоимость жилья в Гуанчжоу, включая ориентацию окна, вид на зеленые насаждения, высоту пола, близость к лесным участкам и водоемам, а также воздействие шума от автомобилей. В работе использовались данные о цене сделки и структурные характеристики 652 жилых единиц в четырёх частных жилых комплексах. Были построены две функциональные модели метода гедонического ценообразования, линейная и полулогарифмическая. Предлагаемая полулогарифмическая модель дала сравнительно более достоверную оценку.

В оценке гедонистических моделей цен на жилье, потенциальный аспект измерений по загрязнению воздуха часто игнорируется. В этой статье ¹² оценивается степень влияния эффекта улучшения качества воздуха на итоговые эмпирические оценки предельной готовности платить (MWTP) на обширной выборке из более чем 100 000 индивидуальных продаж домов в южной Калифорнии. Авторы учитывают пространственную зависимость и эндогенность с использованием недавно разработанных методов пространственной оценки 2SLS. Для пространственной автокорреляции и гетероскедастичности в терминах ошибок используется оценка Келеджийско-Пручинского.

В работе¹³ гедонистический метод ценообразования был использован для анализа влияния структурных переменных, включая цену сделки с землей, расстояние до центра города, центрального делового района, железнодорожного вокзала и больницы, коэффициент площади, количество автобусных линий поблизости и дихотомические переменные, в том числе близость к железнодорожному транзиту, рекреационные объекты и парки, что отражает доступность и условия жизни, на цену сделки с жильем. Были построены гедонистические модели ценообразования, включая модель линейной и полулогарифмической регрессии.

Целью данного исследования¹⁴ является предложение инновационного подхода к оценке недвижимости, называемого количественным сравнительным подходом, для преодоления недостатков субъективных решений по поправочным коэффициентам традиционного сравнительного подхода на основе линейной регрессии.

¹¹ Jim, C. Y., & Chen, W. Y. (2006). Impacts of urban environmental elements on residential housing prices in Guangzhou (China). *Landscape and Urban Planning*, 78(4), pp. 422-434.

¹² Anselin, L., Lozano-Gracia N. Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality, *Empirical Economics* 34 (1), 2008: pp. 5-34.

¹³ He Chengjie, Zhen Wang, Huaicheng Guo, Hu Sheng, Rui Zhou, Yonghui Yang. Driving Forces Analysis for Residential Housing Price in Beijing. *Procedia Environmental Sciences*, 2010: pp. 925 – 936.

¹⁴ I-Cheng Yeh, Tzu-Kuang Hsu. Building real estate valuation models with comparative approach through case-based reasoning. *Appl. Soft Comput.* 65: pp. 260-271 (2018)

В работе¹⁵ оценка внешних воздействий окружающей среды на цены на недвижимость также производится с помощью линейной регрессии.

Авторы этой статьи¹⁶ разработали модель гедонической цены на жилье с использованием объясняющих переменных из четырех измерений, а именно, структуры, района, местоположения и ландшафта, чтобы количественно оценить влияние различных ландшафтов на цену жилья в городе Ханчжоу.

В статье¹⁷ предлагается модель гедонистической регрессии для оценки цен на жилье и пространственной изменчивости цен в течение нескольких лет. Сравниваются два метода: регрессия-коклинг (RCK) и модель коклинг с внешнимд рейфом (CKED), также известная как универсальный коклинг (UCK).

Авторы в своей работе¹⁸ используют географически взвешенную регрессию с временными эффектами.

В данной работе¹⁹ рассматривается долгосрочное влияние и краткосрочная динамика макроэкономических переменных по международным ценам на жилье. Авторы применяют коинтеграционный анализ панельных данных, состоящих из 15 стран в течение 30 лет. Объединение наблюдений позволяет преодолеть ограничения данных, с которыми сталкиваются исследователи при тестировании долгосрочных отношений среди отдельных временных рядов недвижимости.

Работы, основанные на индексах цен.

Индексы цен представляют собой индикаторы для оценки усреднённой цены на квартиры в определённый момент времени, рассчитанные экспертами разными методами.

В работе²⁰ используются данные о почти миллионе домов, проданных в четырех мегаполисах - Атланте, Чикаго, Далласе и Сан-Франциско - для построения квартальных индексов существующих цен на жилье в период с 1970 по 1986 годы. Авторы предлагают и применяют новый метод построения таких индексов, который они называют методом взвешенных повторных продаж (WRS). Результаты по мнению авторов дают точную картину фактического роста цен на жилье в четырех городах. В статье объясняется метод построения

¹⁵ Del Giudice V., De Paola P., Manganelli B., Forte F. The Monetary Valuation of Environmental Externalities through the Analysis of Real Estate Prices. Sustainability 2017, 9, 229.

¹⁶ Haizhen Wen, Yan Qin Zhang, Ling Zhang. Assessing amenity effects of urban landscapes on housing price in Hangzhou, China. Urban Forestry & Urban Greening 14(4), 2015: pp. 1017-1026.

¹⁷ Chica-Olmo J., Cano-Guervos R., Chica-Rivas M. Estimation of Housing Price Variations Using Spatio-Temporal Data. Sustainability 11(6), 2019: pp. 1-21.

¹⁸ Bo Huang, Bo Wu, Michael Barry, Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. International Journal of Geographical Information Science 24(3). 2010: pp. 383-401.

¹⁹ Adams Z., Füss R. Macroeconomic Determinants of International Housing Markets. Journal of Housing Economics 19(1). 2010: pp. 38-50.

²⁰ Case, K. E., Shiller R. J. Prices of Single Family Homes Since 1970: New Indexes for Four Cities, New England Economic Review (1987), pp. 45-56.

индекса, обсуждаются результаты и сравниваются с данными Национальной ассоциации риэлторов о средней цене существующих домов на одну семью за период 1981-1986 годов. В работе также исследуются влияние личного дохода на цену жилья.

В следующей работе²¹ сравниваются результаты с использованием традиционной модели взвешенных повторных продаж (WRS) с результатами, полученными от гибридной модели на основе всей доступной информации о продажах домов. Результаты показывают важность временной дезагрегации в оценке цен на жилье и волатильности, независимо от выбранной модели.

Автор в своей работе²² сравнивает четыре традиционных показателя повторных продаж с недавно разработанным авторегрессионным индексом, который использует методологию повторных продаж и также включает в себя параметр отдельных продаж и эффект местоположения. В статье проведены качественные сравнения по статистическим вопросам, включая влияние времени разрыва в продажах, использование гедонистического подхода, а также обработку разовых и повторных продаж. Кроме того, прогнозирующая способность используется в качестве количественного показателя в анализе с использованием данных о продажах домов в двадцати городских районах США.

Автор статьи²³ описывает так называемый индекс настроения покупки жилья на основе условий продажи, доходов, проблем с работой, цены дома и ожидаемой ставки по ипотечному кредиту.

Эта работа²⁴ направлена на изучение роли рыночных настроений на рынке жилья. Индекс настроения разработан авторами, чтобы отразить совокупное поведение инвесторов и, таким образом, определить варианты сценариев на рынке.

Авторы в своей работе²⁵ применяют регрессию для разработки индексов цен.

Нерегрессионные модели оценки недвижимости.

В статье²⁶ авторы используют методику искусственных нейронных сетей наряду с байесовским подходом на выборке недвижимости. Распределение выходных данных было рассчитано с помощью численного интегрирования в пространстве весов с использованием

²¹ Englund P., Quigley J.M., Redfearn C.L. The Choice of Methodology for Computing Housing Price Indexes: Comparisons of Temporal Aggregation and Sample Definition. *The Journal of Real Estate Finance and Economics* 19, pp. 91–112(1999)

²² Epley D. Assumptions and restrictions on the use of repeat sales to estimate residential price appreciation. *Journal of Real Estate Literature*. 2016; 24; 2: pp. 275–286.

²³ Wilcox J.A. The Home Purchase Sentiment Index: A New Housing Indicator. *Business Economics* 50, pp. 178–190(2015).

²⁴ Chi-man Hui E., Ziyong Wang. 2014. Market sentiment in private housing market. *Habitat International*, 44: pp. 375–385.

²⁵ Martin J.B., Muth R.F., Nourse H.O. A Regression Method for Real Estate. Price Index Construction, *Journal of the American Statistical Association*, 58:304 (1963), pp. 933–942.

²⁶ Peterson S., Flanagan A.B. Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research, American Real Estate Society*, 31(2). 2009: pp. 147–164.

метода цепей Маркова с использованием гибридного метода Монте-Карло (MCHMCM). В той же выборке недвижимости этот подход сравнивался с моделью нейронных сетей (NN), традиционным множественным регрессионным анализом (MRA) и полупараметрическим методом с применением сплайн-штрафов (PSSM). Все четыре метода были применены и оценены для проверки прогнозирующей способности и надежности MCHMCM в сфере недвижимости. Гибридный метод Монте-Карло с цепью Маркова оказался лучшей моделью с абсолютной средней процентной погрешностью 6,61%.

В этой работе²⁷ впервые используются случайный лес в качестве потенциальной техники для массовой оценки жилой недвижимости. В эмпирическом исследовании с использованием данных по жилым квартирам метод показал лучшие результаты, чем такие методики, как CHAID, CART, KNN, множественный регрессионный анализ, искусственные нейронные сети (MLP и RBF) и boosted trees. Также в работе представлен подход для автоматического обнаружения сегментов с систематически завышенными или заниженными результатами прогнозирования. Этот подход применим к различным экспертным системам, которые используются для массовой оценки недвижимости.

В следующей статье²⁸ авторы сравнивают традиционные подходы к оценке объектов недвижимости, основанные на использовании линейной регрессии, методе наименьших квадратов, с методами вычислительного интеллекта: support vector machine (SVM) regression, multi layer perceptron (MLP) и committee of predictors.

Самоорганизующаяся карта (SOM), генерирующая кластеры зон значений, использовалась для получения зависимых от данных весов агрегации. Экспериментальные исследования, проведенные с использованием данных, предоставленных Регистровым центром Литвы, показали очень многообещающие результаты. Производительность методов, основанных на вычислительном интеллекте, была значительно выше тех, которые были получены с использованием традиционных моделей оценки недвижимости Регистрационного центра. Результаты работы комитета с использованием весов, основанных на зонах, полученных из SOM, также были выше, чем в случае использования зон стоимости недвижимости, предоставленных Центром регистрации.

Авторы в своём исследовании²⁹ предлагают новый гибридный метод прогнозирования, который объединяет метод разложения сигналов в функции, «эмпирические моды» (EEMD)

²⁷ Antipov E., Pokryshevskaya E. (2010). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications* 39(2). 2012: pp. 1772-1778.

²⁸ Kontrimas V., Verikas A.A. The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing* 11(1). 2011: pp. 443-448.

²⁹ Plakandaras V., Gupta R., Gogas P., Papadimitriou T. Forecasting the U.S. Real House Price Index. *Economic Modelling*, Elsevier, 45(C). 2015: pp. 259-267.

из области обработки сигналов с методом опорных векторов (SVR). Авторы тестируют способность прогнозирования предлагаемой модели по сравнению с моделью случайного блуждания (RW), байесовской авторегрессионной моделью и байесовской векторной моделью авторегрессии.

В этой работе авторы³⁰ используют оценку с помощью индекса цен и модель авторегрессии с плавным переходом (STAR).

Здесь авторы предлагают улучшенный метод для определения динамики цены во времени на уровне каждого дома по сравнению с традиционными индексами цен на жильё, такими как индекс Case-Shiller.³¹ Используются градиентное усиление (Gradient Boosting), деревья регрессии (Regression Trees), машинное обучение, сглаживающий сплайн (Spline Smoothing).

В этом исследовании³² оценивается влияние визуальных характеристик дома на его рыночную цену, используя сверточные нейронные сети на большом наборе фотографий домашних интерьеров и экстерьеров для оценки уровня роскоши фотографий недвижимости. Также авторы разрабатывают новую структуру для автоматизированной оценки стоимости с использованием вышеуказанных фотографий в дополнение к характеристикам дома, включая размер, предлагаемую цену и количество спален.

В этой работе³³ применяют нелинейную авторегрессию со скользящим средним с экзогенными входами (NARMAX) в сочетании с векторной моделью коррекции ошибок, чтобы исследовать детерминанты цен на жильё в Китае за период с 1999 до 2010.

В исследовании³⁴ представлен гибридный подход генетического алгоритма и метода опорных векторов (G-SVM) в прогнозировании цен на жильё в Китае.

В исследовании³⁵ авторы в качестве методологии исследования для разработки модели прогнозирования цен на жильё используют алгоритмы машинного обучения, такие как C4.5, RIPPER, Naïve Bayesian и Ada Boost.

В работе³⁶ для прогнозирования цен на недвижимость используется метод опорных векторов, а для определения его параметров метод ряда частиц (PSO).

³⁰ Kouwenberg R., Zwinkels R. Forecasting the US housing market. *International Journal of Forecasting*, 30(3). 2014: pp. 415-425.

³¹ Barry J.R., Ellisy E.A., Kassaby A., Redfearnz C.L., Srinivasany N., Voris K.B. Home Price Index: A Machine Learning Methodology. *International Journal of Semantic Computing* 1(1), 2017: pp. 9-22.

³² Poursaeed O., Matera T., Belongie S. Vision-based real estate price estimation. *Machine Vision and Applications* 29. 2018: pp. 667–676.

³³ Zhang, Yanbing & Hua, Xiuping & Zhao, Liang. Exploring determinants of housing prices: A case study of Chinese experience in 1999–2010. *Economic Modelling*, Elsevier, vol. 29(6). 2012: pp. 2349-2361.

³⁴ Jirong Gu, Mingcang Zhu, Liuguangyan Jiang. Housing price forecasting based on genetic algorithm and support vector machine. *Expert Syst. Appl.* 38(4): pp. 3383-3386 (2011).

³⁵ Byeonghwa Park, Jae Kwon Bae. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems With Applications* 42(6).2015: pp. 2928-2934.

Использование нечётких множеств.

Отдельно хотелось бы рассмотреть работы, основанные на нечётких моделях, которые также могут эффективно оценивать влияние различных характеристик на цену жилья.

В статье³⁷ авторы исследуют методы нечеткого моделирования для прогнозирования цен на жилые помещений, исходя из таких факторов, как жилая площадь помещений, возраст здания, количество комнат в квартире, этаж, на котором расположена квартира, этажность в здании, а также расстояние до центра города. Предлагаемые методы моделирования опираются на два аспекта: первый (Sparse FIS) - это метод оптимизации весов правил внутри процедуры итеративной оптимизации, ограничивающей количество важных правил; второй (FLEXFIS) - это однопроходный инкрементальный метод, который способен адаптировать сгенерированные модели прогнозирования с новыми записями данных по запросу, а также справиться с онлайн-потокими данными. Полученные нечеткие модели дают интерпретируемые результаты зависимостей в форме лингвистически читаемых правил (ЕСЛИ-ТО условия). Оба подхода сравниваются с современными методами оценки жилых помещений, в том числе с такими как скользящее временное окно, метод ближайших соседей и усреднение.

В этой работе³⁸ используются нечеткие логические системы, учитывающие планы города, близость к культурным, медицинским, образовательным учреждениям, системы общественного транспорта, другие факторы окружающей среды и технологической модернизации. Построенная модель была применена для прогнозирования цен домов, расположенных в разных регионах города Эскисэхир в Турции.

Цены-пузыри.

Ряд работ посвящён такому явлению в экономике, как цены-пузыри, когда цена на рынке недвижимости сильно отличается от справедливой.

Целью исследования³⁹ является изучение существования пузыря цен на жилье в США с помощью модели коррекции ошибок.

Авторы⁴⁰ используют данные по шести столичным рынкам жилья в трех странах, чтобы сравнить методы, используемые для измерения пузырей цен на жилье. Авторы использовали метод оценки активов для ретроспективной идентификации периодов пузырей.

³⁶ Xibin Wang, Junhao Wen, Yihao Zhang, Yubiao Wang. Real estate price forecasting based on SVM optimized by PSO. *Optik*, 125(3). 2014: pp. 1439-1443.

³⁷ Lughofer E., Trawinski B., Trawinski K., Kempa O., Lasota T. On employing fuzzy modeling algorithms for the valuation of residential premises. *Information Science* 181(23). 2011: pp. 5123-5142.

³⁸ Hakan Kusan, Osman Aytekin, Ilker Özdemir: The use of fuzzy logic in predicting house selling price. *Expert Syst. Appl.* 37(3). 2010: pp.1808-1813.

³⁹ Clark S.P., Coggin, T. D. Was there a U.S. house price bubble? An econometric analysis using national and regional panel data. *The Quarterly Review of Economics and Finance*, Elsevier, 51(2). 2011: pp. 189-200.

В работах ^{41 42 43 44} применяются разные варианты авторегрессии для оценки данного явления.

В последнее время появилось множество методов оценки объектов недвижимости, в том числе и на больших данных, что упрощает анализ и исследование в данной работе.

⁴⁰ Bourassa S.C., Hoesli M., Oikarinen E. Measuring House Price Bubbles," Real Estate Economics, American Real Estate and Urban Economics Association, 47(2). 2019: pp. 534-563.

⁴¹ Fabozzi, F. J., Xiao, K. The Timeline Estimation of Bubbles: The Case of Real Estate. Real Estate Economics 47(2). 2019: pp. 564-594.

⁴² Fernandez-Kranz, D., Hon, M.T. A cross-section analysis of the income elasticity of housing demand in Spain: Is there a real estate bubble? Journal of real estate finance and economics. 32(4). 2006: pp. 449-470.

⁴³ Phillips, P.C.B., Shi, S.-P. & Yu, J. Testing for Multiple Bubbles: Historical Episodes of Exuberance. International Economic Review 56(4). 2015: pp. 1043– 1078.

⁴⁴ Phillips, P.C.B., Shi, S.-P., & Yu, J. Testing for Multiple Bubbles: Limit Theory of Real Time Detectors. International Economic Review 56(4). 2015: pp. 1079– 1134.

ГЛАВА 2. АНАЛИЗ НА ОСНОВЕ СРАВНЕНИЯ РЫНОЧНОЙ И КАДАСТРОВОЙ СТОИМОСТИ ОБЪЕКТОВ

2.1. Наиболее вероятная траектория рыночной стоимости во времени Данные

Данные для исследования представляют собой два массива – рыночные и кадастровые.

Рыночные данные взяты из источника «Бюллетень недвижимости Санкт-Петербурга». Их можно найти в печатном издании и на портале bn.ru. Бюллетень выходит по четвергам 1 раз в неделю. Каждому выпуску присваивается номер, например номер № 1531 — это номер от 17 августа 2012 года. Данные в массиве представлены с августа 2012 года по декабрь 2018 года.⁴⁵

Кадастровые данные также представлены двумя массивами - за 2015 и за 2018 год. Данные за 2015 год есть в открытом доступе на сайте Росреестра. Данные за 2018 год на сайте ГБУ «Кадастровая оценка» Санкт-Петербурга.^{46,47}

Для того чтобы начать работать с данными, необходимо для начала определить к какому типу они относятся – большим или традиционным.

Согласно сайту Oracle, крупнейшему поставщику программного обеспечения для организаций, базовое определение большим данным дала компания Gartner в 2011 году: «Большие данные — это разнообразные данные, которые поступают с постоянно растущей скоростью и объем которых постоянно растет». Основные свойства больших данных, которые можно выделить:

- Большой размер, с которым не справляются традиционные программы для обработки данных
- Разнообразие – помимо чисел и текста большие данные также могут включать картинки, аудио файлы, видео файлы и другие возможные типы данных. Большие данные обычно не структурированы или частично структурированы.
- Высокая скорость поступления. Для больших данных также характерно извлечение в реальном времени.

Для данных, которые используются в данном исследовании не выполняются все условия соответствия определению, но, как было сказано ранее, термин большие данные

⁴⁵ URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости

⁴⁶ Отчет об определении кадастровой стоимости объектов недвижимости (за исключением земельных участков), расположенных на территории Санкт-Петербурга. КУГИ Правительства Санкт-Петербурга, том 2, раздел 2.3, СПб, 2012 http://rosreestr.ru/wps/portal/p/cc_ib_portal_services/cc_ib_ais_fdg/co/ Субъект РФ: С-Петербург. Виды объектов: Недвижимость, помещения. Отчет №32-1-0733/2012(2), Санкт-Петербург, 30.11.2012

⁴⁷ Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 г. <http://guko.commim.gov.spb.ru/>

употребляется в данном контексте в том смысле, что их объем, не позволяет без предварительной подготовки и модификаций применять традиционные статистические методы.⁴⁸

Поскольку во всеобщем понимании данные относятся к традиционным, сбор данных, их очистка, решение проблем с пропущенными значениями и отнесение тех или иных данных к разным типам - категориальным и количественным, было произведено с помощью стандартных программ для обработки традиционных данных – Excel. Также анализ таких данных может производиться с помощью языков программирования R, Python, SQL и Matlab.

После того как данные были собраны и организованы, необходимо извлечь из них информацию, чтобы получить представление о данных с точки зрения рынка и бизнеса. Для этого можно использовать приёмы и инструменты Business Intelligence аналитики. Эта область с помощью визуализаций и метрик помогает выделить основные паттерны в данных и наметить дальнейшие направления для продвинутой аналитики. Языки программирования в этой области те же, что и для обработки традиционных типов данных. Помимо Excel также используется такое программное обеспечение, как Power BI, Tableau, SaaS и Qlik. В данном случае использовался Power BI в силу его доступности.

Первым этапом было соединение массива кадастровых и рыночных данных. Кадастровые данные имеют уникальное поле – кадастровый номер, поэтому их достаточно легко структурировать. Все остальные поля с данными имеют чёткий и единообразный вид. Рыночные данные не имеют привязки к кадастровому номеру и не имеют уникального поля. Кроме того, текстовое поле с адресом имеют отличный от кадастровых данных вид – присутствуют сокращения, отсутствуют знаки препинания и связующие слова, такие как «дом».

Было принято решение соединить массивы по созданному вручную ключевому полю, которое объединяет такие понятия, как адрес и общая площадь объекта, потому что эти характеристики жилой недвижимости не меняются. В среде Power BI была разработана функция, которая создаёт шаблон для обработки файла с рыночными данными, преобразуя адрес в близкий к кадастровому вид и объединяет поля для ключевого поля с необходимыми преобразованиями в типах данных. Результат представлен на рисунке 2.1. Слева названия уже загруженных файлов. В пустое поле filepath необходимо ввести путь к файлу на персональном компьютере. После введения функция автоматически преобразует файл, в котором будет изменён адрес по шаблону близкому к адресу в кадастровой базе, а также будут соединены необходимые поля для получения уникального ключевого поля.

⁴⁸ URL: <https://www.oracle.com/ru/big-data/guide/what-is-big-data.html#close> – официальный сайт Oracle в России и СНГ

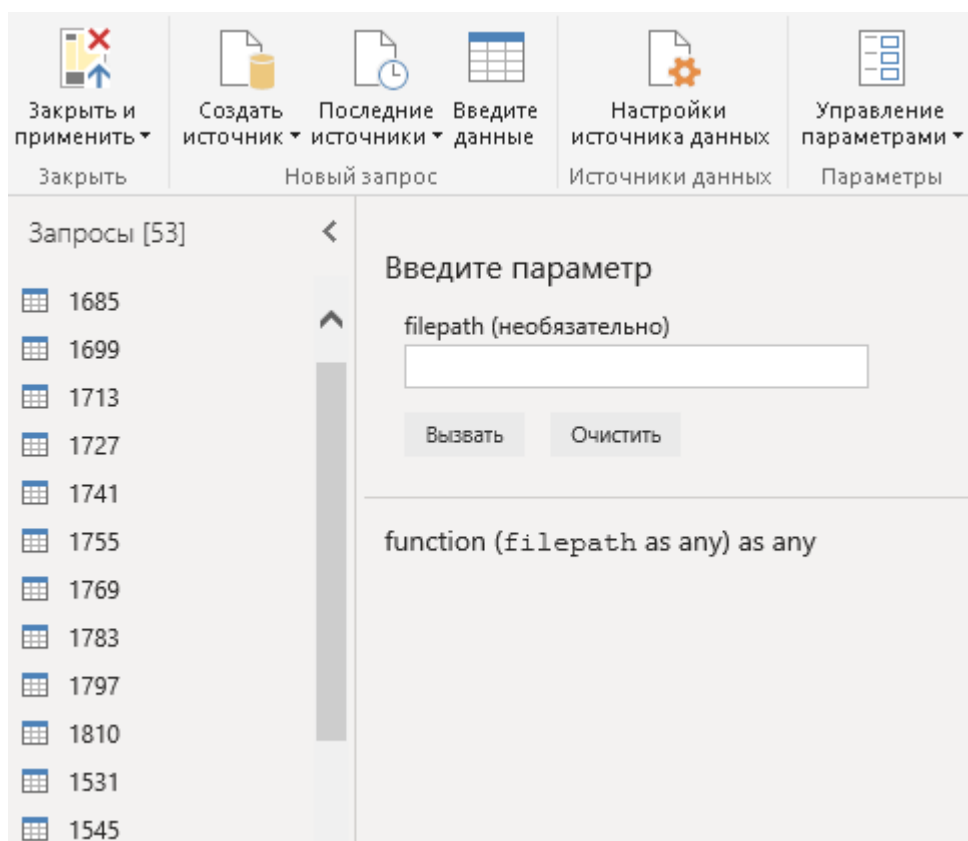


Рис. 2.1. Вид преобразовывающей функции в расширенном редакторе Power BI

Составлено по: собственное исследование автора

Примечание: **filepath** – поле для введения пути к файлу на персональном компьютере.

1685, 1699, 1713 и т.д. – названия загруженных с помощью разработанной функции файлов, которые соответствуют номерам в журнале Бюллетень недвижимости.

Затем эта функция применялась для всего массива файлов с рыночными ценами с промежутком в 14 номеров или 1 квартал. Все множество связей представлено на рисунке 2.2. На основе полученных пар кадастровой и рыночной стоимостей были построены необходимые визуализации и произведены вычисления, которые будут описаны ниже в ходе выполнения основной работы.

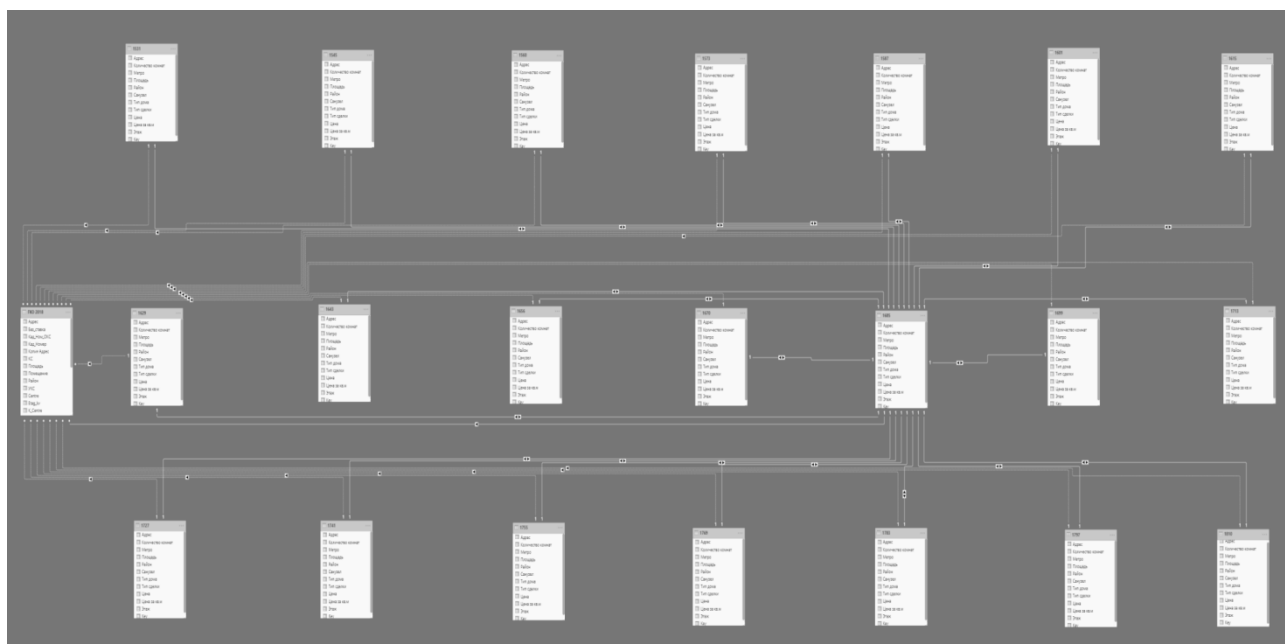


Рисунок 2.2. Множество файлов с установленными связями по искусственно созданному ключевому полю

Составлено по: собственное исследование автора

Все вышеуказанные действия с данными относятся к прошлому, они обрабатывают массивы операций, которые завершились и имеют результат в настоящем. Для того, чтобы анализировать будущее или делать прогнозы используются инструменты области Data Science. Здесь также есть два раздела – традиционные методы и машинное обучение (Machine Learning – ML). Обычно традиционные методы применяются к объёмам данных, которые можно отнести к традиционным, а методы машинного обучения применяются к большим данным.

К традиционным методам относятся кластерный анализ, факторный анализ, регрессионный анализ, логистическая регрессия, временные ряды и другие. Для традиционных методов используются R, Python, Matlab и такое программное обеспечение, как Excel, IBMSPSS, Stata и EViews. Именно с таких методов начнётся дальнейший анализ и моделирование прогнозов в данной работе.

Методы машинного обучения можно разделить на обучение с учителем (supervised learning) и без учителя (unsupervised learning). Существуют еще методы обучения с подкреплением (reinforcement learning), они похожи на обучение с учителем, но целью является не минимизация потерь, а максимизация выигрыша, целевой функции. К таким методам относятся нейронные сети (Neural Networks), метод опорных векторов (SVM), метод k-means, глубокое обучение (Deep Learning), случайный лес (Random Forest) и другие. Здесь помимо вышеуказанных языков программирования используются такие высокоуровневые

языки как C, C++, Java, JavaScript и Scala. Из программного обеспечения можно выделить платформу RapidMiner и Microsoft Azure.

Далее необходимо обратиться к анализу с точки зрения математики и статистики для обоснования выдвигаемых гипотез.

Общая формула

Для отслеживания рыночной стоимости как наиболее вероятной цены для каждого конкретного кластера объектов недвижимости (под кластером понимается некоторое подмножество всех объектов представленных в кадастровых базах, удовлетворяющее определенному набору качественных ценообразующих факторов) необходимо рассмотреть совместные распределения кадастровых стоимостей и рыночных данных за некоторый интервал времени. Затем по кадастровому номеру определить кадастровую стоимость и определить наиболее вероятный ансамбль рыночных стоимостей на всем интервале наблюдений.

Пусть V_{KC} – кадастровая стоимость, $V(t_1), \dots, V(t_n)$ – рыночная цена на моменты времени t_1, \dots, t_n . $V_{KC}, V(t_1), \dots, V(t_n)$ предполагаются логарифмически нормальными случайными величинами. Пусть $W = \ln(V_{KC}), Y_i = \ln(V(t_i)), i = 1, n$ (тогда $V_{KC} = e^W, V(t_i) = e^{Y_i}$).

Рассмотрим многомерный нормальный случайный вектор (W, Y_1, \dots, Y_n) с вектором Средних $(\mu_W, \mu_{Y_1}, \dots, \mu_{Y_n})$. Ковариационную матрицу запишем в блочном виде:

$$CV = \begin{pmatrix} \sigma_W^2 & cov(W, \vec{Y}) \\ cov(W, \vec{Y})^T & COV \end{pmatrix} \quad (1),$$

где COV – ковариационная матрица случайного вектора $\vec{Y} = (Y_1, \dots, Y_n)$, под $cov(W, \vec{Y})$ понимается вектор $cov(W, \vec{Y}) = (\rho_{WY_1} \sigma_W \sigma_{Y_1}, \dots, \rho_{WY_n} \sigma_W \sigma_{Y_n})$, $\sigma_W^2, \sigma_{Y_1}^2, \dots, \sigma_{Y_n}^2$ – дисперсии случайных величин W, Y_1, \dots, Y_n , $\rho_{WY_1}, \dots, \rho_{WY_n}$ – соответствующие коэффициенты корреляции.

Условное математическое ожидание вектора $\vec{Y} = (Y_1, \dots, Y_n)$, при условии, что $W = w$

$$E(\vec{Y} | W = w) = \vec{\mu} + \frac{cov(W, \vec{Y})^T}{\sigma_W^2} (w - \mu_W), \quad (2)$$

где $\vec{\mu} = (\mu_{Y_1}, \dots, \mu_{Y_n})$.

условная ковариационная матрица при условии, что $W = w$

$$COV(\vec{Y} | W = w) = COV - \frac{cov(W, \vec{Y})^T \times cov(W, \vec{Y})}{\sigma_W^2} \quad (3)$$

В соответствии с формулировкой ФЗ-135 [1] под рыночной стоимостью понимается наиболее вероятная цена, по которой объект оценки может быть отчужден на открытом

рынке в условиях совершенной конкуренции. В контексте такого понимания рыночной стоимости будем искать наиболее вероятное значение вектора $V(t_1), \dots, V(t_n)$, при условии, что кадастровая стоимость фиксирована $V_{KC} = v_{KC}$.

Утверждение

Абсолютный максимум (мода) плотности случайного логарифмически нормального вектора \vec{x} достигается в точке с координатами $\exp(\vec{\mu} - \Sigma \times \mathbf{1})$, где $\vec{\mu}$ - вектор математических ожиданий логарифмов компонент, Σ - ковариационная матрица логарифмов компонент, $\mathbf{1}$ - вектор, состоящий из единиц. Доказательство представлено в статье [2].

В данном случае наиболее вероятная траектория рыночных цен во времени для каждого заданного значения кадастровой стоимости $V_{KC} = v_{KC}$ определяется по формуле:

$$MODE(V(t)|V_{KC} = v_{KC}) = \exp(E(\vec{Y}|W = w) - COV(\vec{Y}|W = w) \times \mathbf{1}) \quad (4)$$

($\mathbf{1}$ - вектор, состоящий из единиц).

Расчёт для двух моментов времени

Так как невозможно представить визуализацию совместных распределений размерности больше трех, рассмотрим сначала трехмерный случай. Для численного примера выбраны значения кадастровых стоимостей и данные бюллетеня недвижимости по Приморскому району г. Санкт-Петербурга N1685 (2 квартал 2015 года), N1699 (3 квартал 2015 года). Выбор района объясняется тем, что это самый крупный по численности населения район г. Санкт-Петербурга (более 568 516 человек) характеризуется, в основном, современной застройкой и застройкой позднего советского периода. Рассматриваются распределения кадастровой стоимости 1 квадратного метра жилой площади (в отдельных квартирах), рыночных данных на 2 и 3 кварталы 2015 года⁴⁹. Важно отметить, что дата определения кадастровой стоимости не имеет значения - важно, что она установлена и занесена в кадастровую базу данных.

Для распределения кадастровых стоимостей по Приморскому району по кадастровой базе 2018 года получены следующие значения параметров логарифмически нормального распределения: $\mu_{KC} = 4,63$, $\sigma_{KC} = 0,12$. Для выборки рыночных данных из номера БН1685 получены значения параметров логарифмически нормального распределения: $\mu_{1685} = 4,656$, $\sigma_{1685} = 0,17$, для выборки рыночных данных из номера БН1699 получены значения параметров логарифмически нормального распределения: $\mu_{1699} = 4,65$, $\sigma_{1699} = 0,18$, получены также выборочные коэффициенты корреляции $\rho_{KC,1685} = 0,52$, $\rho_{KC,1699} = 0,62$, $\rho_{1685,1699} = 0,6904$.

⁴⁹ Данные кадастровая база г. Санкт-Петербурга 2018 года, данные Бюллетеня недвижимости N1685,1699.

Проверка на совместную нормальность логарифмов кадастровой стоимости и рыночных данных

В условиях развивающегося рынка и общей доступности информации по ценам и объектам предложения на рынке недвижимости применение статистических методов позволяет профессиональным участникам рынка определить цену, по которой с высокой степенью вероятности может состояться сделка на дату оценки. Для корректного применения многих статистических методов требуется предварительная проверка гипотезы о том, что выборочные данные принадлежат генеральной совокупности, имеющей нормальное распределение. Это одно из требований и к применению методов, использующихся в этом исследовании. Особенностью сопоставления данных кадастровой базы и рыночных данных является неполнота рыночных данных. Рыночные данные, как правило, плохо структурированы, содержат большое количество ошибок, не корректных записей, не правильно заполненных полей (с точки зрения анализа данных). Кадастровая база, наоборот, хорошо структурирована, единообразна, содержит все существующие объекты, прошедшие кадастровый учет, ошибки в определении кадастровой стоимости не являются доминирующими. Для сравнения кадастровой базы, рыночных данных использовался разработанный авторами ключ, учитывающий неизменные характеристики объектов: адрес, номер помещения, площадь.

Для тестирования полученных выборок использовался параметрический критерий Колмогорова-Смирнова.

Проверка маргинальных распределений

На рисунке 2.3. показана гистограмма и кривая плотности нормального распределения для логарифмов цен из номера БН №1685.

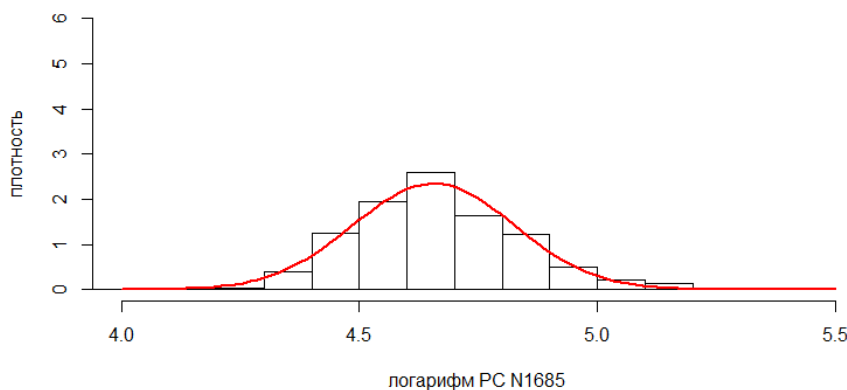


Рис.2.3. Гистограмма и кривая плотности нормального распределения для логарифмов цен из БН №1685

Составлено по: [URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости

Примечание: РС – рыночная стоимость

Объем выборки 469 элементов, значение p-value теста Колмогорова-Смирнова $p\text{-value} = 0.3617$.

На рисунке 2.4 показана гистограмма и кривая плотности нормального распределения для логарифмов цен из номера БН №1699.

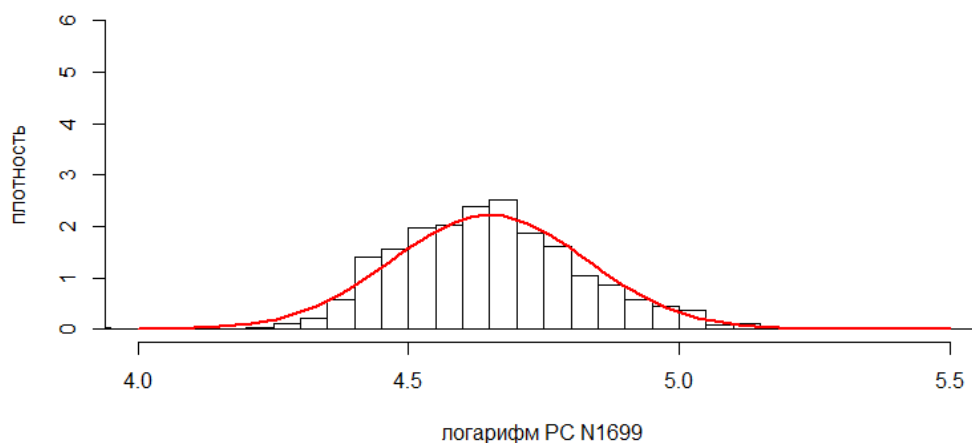


Рис.2.4. Гистограмма и кривая плотности нормального распределения для логарифмов цен из БН №1699

Составлено по: [URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости

Примечание: РС – рыночная стоимость

Объем выборки 484 элемента, значение p-value теста Колмогорова-Смирнова $p\text{-value} = 0.403$.

На рисунке 2.5 показана гистограмма и кривая плотности нормального распределения для логарифмов кадастровых цен по Приморскому району.

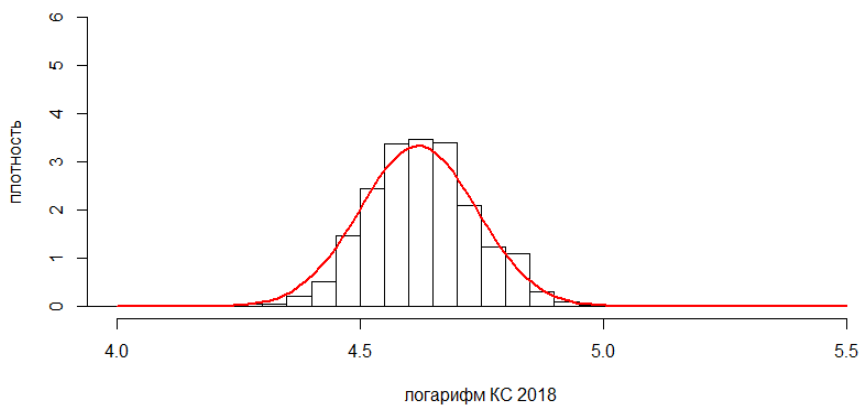


Рис.2.5. Гистограмма и кривая плотности нормального распределения для логарифмов кадастровых цен по Приморскому району

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

Примечание: **КС** – кадастровая стоимость

Первоначальный объем выборки 265 950 элементов, после удаления крайних значений в выборке осталось 264 687 элементов (99,52%). Выборка остается достаточно большой для статистических тестов. Как правило, тесты на соответствие эмпирического и теоретического распределений при объемах выборки более 1000 не работают. Для тестирования такой выборки применялся следующий прием:

- из выборки случайным образом отбиралась подвыборка объемом 500 элементов
- подвыборка тестировалась тестом Колмогорова-Смирнова, результат p-value записывался в отдельный массив
- тест повторялся 1000 раз
- для полученного массива вычислялись минимальные и максимальные значения, общий результат представлен гистограммой.

На рисунке 2.6. показана гистограмма значений p-value теста Колмогорова-Смирнова для 1000 повторений теста.

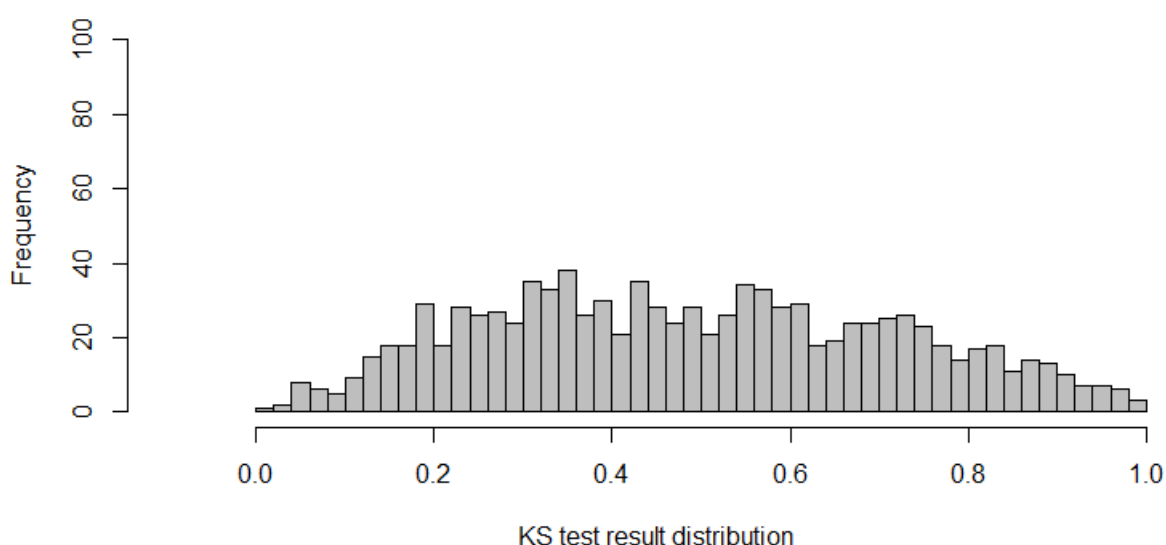


Рис.2.6. Гистограмма значений p-value теста Колмогорова-Смирнова для 1000 повторений теста по случайным выборкам объема 500

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

Примечание: **Frequency** – частота.

KS test result distribution – распределение результатов теста Колмогорова-Смирнова.

Минимальное значение $p\text{-value}=0.0184$, максимальное значение $p\text{-value}=0.9887$, общее количество элементов выборки со значением более 0.05 – 994

Известна следующая теорема: для того, чтобы компоненты случайного вектора были распределены совместно нормально необходимо и достаточно, чтобы любая линейная комбинация компонент была нормально распределена. На основании этого утверждения была построена проверка совместной нормальности кадастровой стоимости по Приморскому району и рыночных данных по номерам БН № 1685 и №1699. Для этого

- случайным образом генерировались весовые коэффициенты, неотрицательные, в сумме равные 1,
- линейная комбинация логарифмов кадастровых стоимостей и рыночных данных тестировалась на нормальность тестом Колмогорова-Смирнова, результат $p\text{-value}$ записывался в отдельный массив
- тест повторялся 1000 раз
- для полученного массива вычислялись минимальные и максимальные значения, общий результат представлен гистограммой.

На рисунках 2.7., 2.8. показаны гистограммы значений $p\text{-value}$ теста Колмогорова-Смирнова для 1000 повторений теста для проверки на совместную нормальность двумерных распределений кадастровой стоимости и рыночных данных.

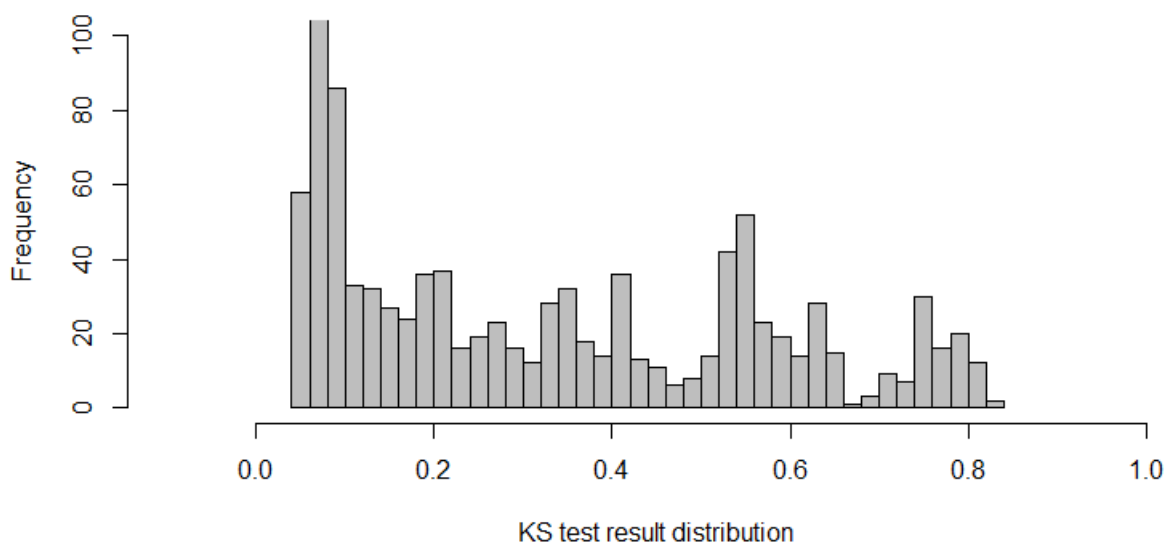


Рис.2.7. Гистограмма значений p-value теста Колмогорова-Смирнова для 1000 повторений теста на совместную нормальность кадастровых стоимостей и рыночных данных из номера БН №1685

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

Примечание: **Frequency** – частота.

KS test result distribution – распределение результатов теста Колмогорова-Смирнова.

Минимальное значение $p\text{-value}=0.0539$, максимальное значение $p\text{-value}=0.8269$. общее количество элементов выборки со значением более 0.05 – 1000.

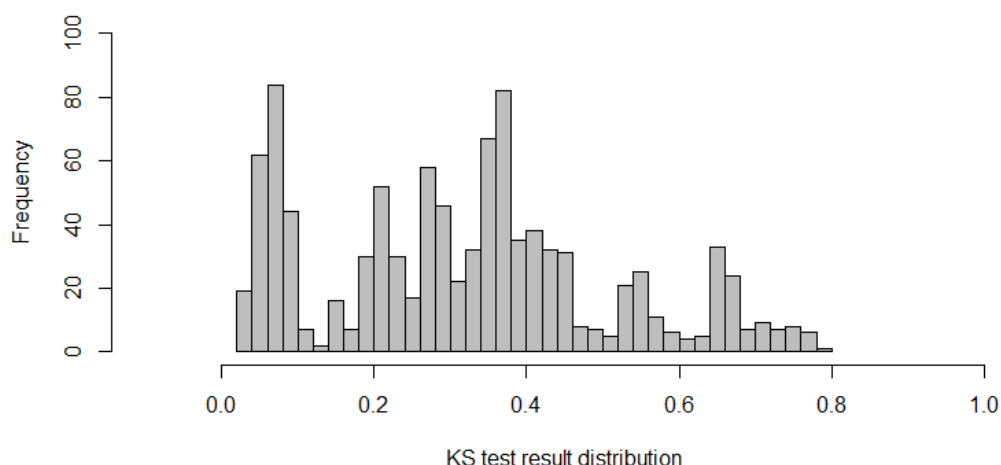


Рис.2.8. Гистограмма значений p-value теста Колмогорова-Смирнова для 1000 повторений теста на совместную нормальность кадастровых стоимостей и рыночных данных из номера БН №1699

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

Примечание: **Frequency** – частота.

KS test result distribution – распределение результатов теста Колмогорова-Смирнова.

Минимальное значение $p\text{-value}=0.0280$, максимальное значение $p\text{-value}=0.7811$, общее количество элементов выборки со значением более 0.05 – 945.

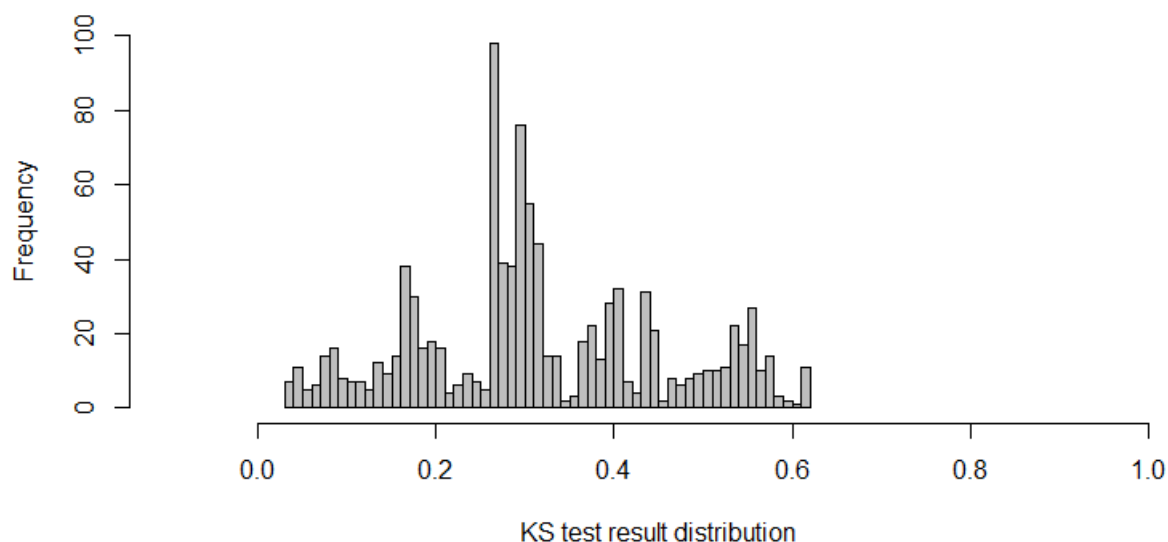


Рис.2.9. Гистограмма значений p-value теста Колмогорова-Смирнова для 1000 повторений теста на совместную нормальность рыночных данных из номера БН №1685 и рыночных данных из номера БН №1699

Составлено по: [URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости.

Примечание: **Frequency** – частота.

KS test result distribution – распределение результатов теста Колмогорова-Смирнова.

Минимальное значение $p\text{-value}=0.0312$, максимальное значение $p\text{-value}=0.6170$, общее количество элементов выборки со значением более 0.05 – 982.

Результаты тестов говорят о том, что есть основания сохранить в качестве рабочих гипотезы:

- о логарифмически нормальных распределениях кадастровых стоимостей, рыночных данных из номера БН №1685, рыночных данных из номера БН №1699,
- о совместной нормальности логарифмов кадастровых стоимостей и рыночных данных из номера БН №1685
- о совместной нормальности логарифмов кадастровых стоимостей и рыночных данных из номера БН №1699

Из ранее полученных данных составим вектор средних

$$(\mu_W, \mu_{Y_{1685}}, \mu_{Y_{1699}}) = (4.63, 4.656, 4.65)$$

и ковариационную матрицу

$$\begin{aligned}
CV &= \begin{pmatrix} \sigma_{KC}^2 & \rho_{KC,1685} \times \sigma_{KC} \times \sigma_{1685} & \rho_{KC,1699} \times \sigma_{KC} \times \sigma_{1699} \\ \rho_{KC,1685} \times \sigma_{KC} \times \sigma_{1685} & \sigma_{1685}^2 & \rho_{1685,1699} \times \sigma_{1685} \times \sigma_{1699} \\ \rho_{KC,1699} \times \sigma_{KC} \times \sigma_{1699} & \rho_{1685,1699} \times \sigma_{1685} \times \sigma_{1699} & \sigma_{1699}^2 \end{pmatrix} = \\
&= \begin{pmatrix} 0,12^2 & 0,52 \times 0,12 \times 0,17 & 0,62 \times 0,12 \times 0,18 \\ 0,52 \times 0,12 \times 0,17 & 0,17^2 & 0,6904 \times 0,17 \times 0,18 \\ 0,62 \times 0,12 \times 0,18 & 0,6904 \times 0,17 \times 0,18 & 0,18^2 \end{pmatrix} \\
&= \begin{pmatrix} 0,0144 & 0,0106 & 0,0134 \\ 0,0106 & 0,0289 & 0,0211 \\ 0,0134 & 0,0211 & 0,0324 \end{pmatrix}
\end{aligned}$$

Теперь для любого значения кадастровой стоимости $V_{KC} = v_{KC}$ последовательно применяя формулы (1)-(4) можно указать наиболее вероятную пару рыночных стоимостей на момент времени, соответствующий номеру БН№1685 и номеру БН№1699. Например, при $V_{KC} = 100$ тыс. руб. за кв. м. наиболее вероятной парой будет $V_{1(1685)} = 98,623$ тыс. руб., $V_{2(1699)} = 97,978$ тыс. руб. Оценки, сделанные не по модальным, а по медианным значениям (т.е. рассчитанные по формуле $MODE(V(t)|V_{KC} = v_{KC}) = \exp(E(\bar{Y}|W = w))$), дали бы несколько завышенный результат $V_{1(1685)} = 102,952$ тыс. руб., $V_{2(1699)} = 102,814$ тыс. руб., не отвечающий принципу наибольшей вероятности. К тому же, оценки, сделанные по медиане, имеют целое множество других равновероятных значений.

На рисунках 2.10., 2.11., 2.12. последовательно показаны рассеяние эмпирических точек, модельные распределения, линии уровней модельных распределений для двумерных логарифмически нормальных распределений пар: (кадастровая стоимость - рыночные цены по номеру БН№1685), (кадастровая стоимость - рыночные цены по номеру БН№1699), (рыночные цены по номеру БН№1685 - рыночные цены по номеру БН№1699).

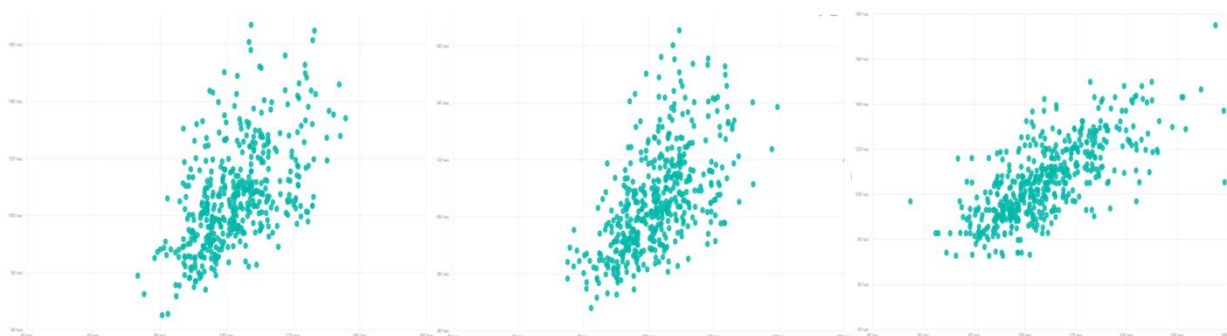


Рис.2.10. Рассеяние исходных выборок

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018

года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

Примечание: слева направо: кадастровая стоимость (по горизонтали)- рыночные цены из номера БН№1685 (по вертикали), кадастровая стоимость (по горизонтали)- рыночные цены из номера БН№1699 (по вертикали), рыночные цены из номера БН №1685(по горизонтали)- рыночные цены из номера БН №1699(вертикали).

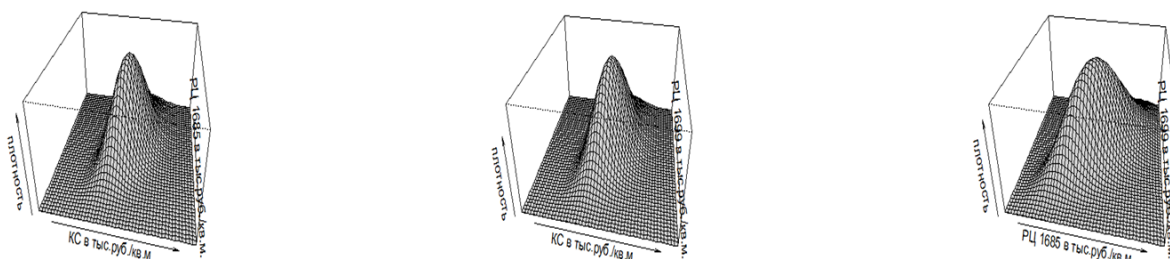


Рис.2.11. Модельные логарифмически нормальные двумерные распределения

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

Примечание: слева направо: кадастровая стоимость - рыночные цены из номера БН №1685, кадастровая стоимость - рыночные цены из номера БН №1699, рыночные цены из номера БН №1685- рыночные цены из номера БН №1699.

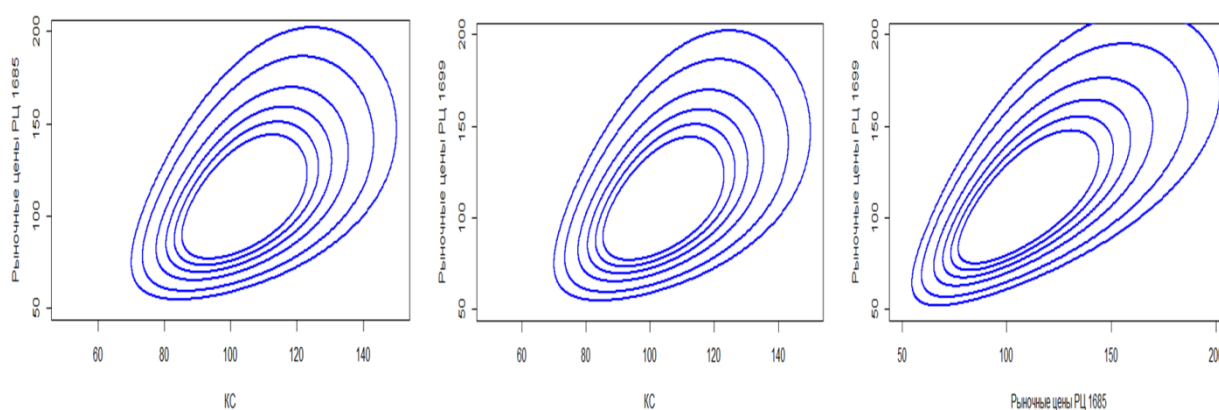


Рис.2.12. Линии уровней равной вероятности для модельных логарифмически нормальных двумерных распределений

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб,

24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

[URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости.

Примечание: слева направо: рыночные цены из номера БН№1685(по вертикали)- кадастровая стоимость (по горизонтали), рыночные цены из номера БН№1699(по вертикали)- кадастровая стоимость (по горизонтали), рыночные цены из номера БН№1685(по горизонтали)- рыночные цены из номера БН№1699(вертикали).

На рисунке 2.11. хорошо видна единственность модальной точки (точки максимума плотности распределения).

На рисунке 2.12. видно, что любая другая точка, кроме модальной (в том числе медианная) имеет множество других точек равной вероятности, с сильно различающимися значениями компонент двумерного вектора.

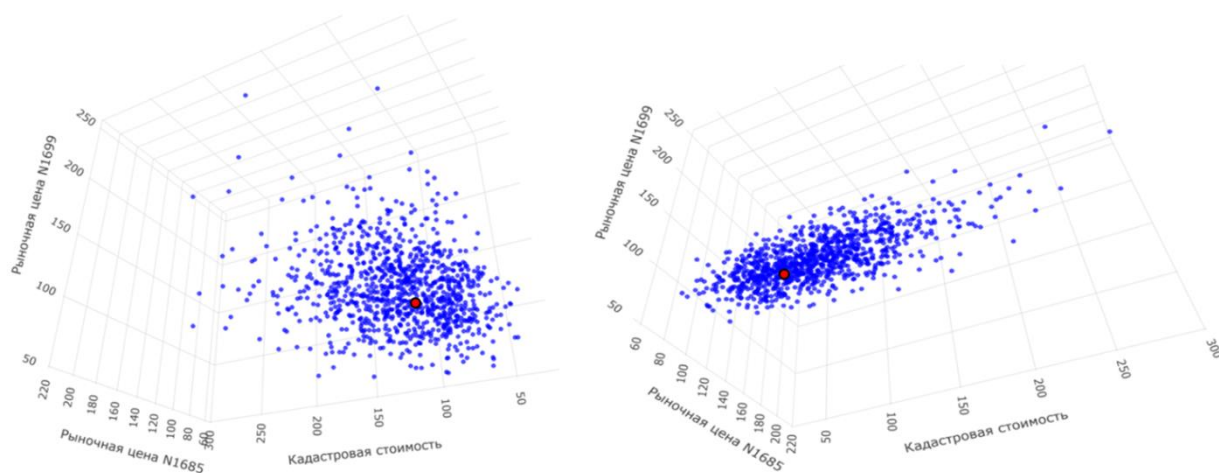


Рис.2.13. Рассеяние (в разных проекциях) трехмерных распределений совместно логарифмически нормальных случайных величин: кадастровая стоимость, рыночные цены из БН№1685, рыночные цены из БН№1699

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

[URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости.

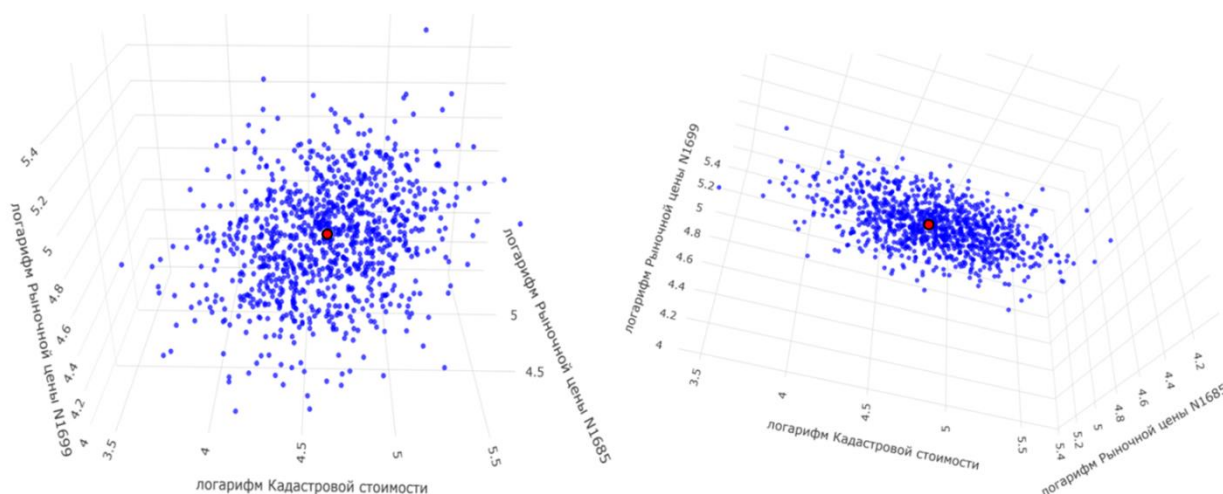


Рис.2.14. Рассеяние (в разных проекциях) трехмерных распределений совместно нормальных случайных величин: логарифм кадастровой стоимости, логарифм рыночных цен из БН№1685, логарифм рыночных цен из БН№1699

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

На рисунке 2.13. в разных проекциях показано рассеяние трехмерных распределений совместно логарифмически нормальных случайных величин: кадастровая стоимость, рыночные цены из БН№1685, рыночные цены из БН№1699. Красная точка расположена ассиметрично и отвечает точке максимальной трехмерной плотности.

На рисунке 2.14. в разных проекциях показано рассеяние трехмерных распределений совместно нормальных случайных величин: логарифм кадастровой стоимости, логарифм рыночных цен из БН№1685, логарифм рыночных цен из БН№1699. Красная точка расположена симметрично и отвечает точке максимальной трехмерной плотности.

Расчет наиболее вероятной траектории рыночных стоимостей для периода 5 лет поквартально (21 момент времени)

Приморский район

Исследованы рыночные данные за период с 4 квартала 2012 по 4 квартал 2017 года по Приморскому району города Санкт-Петербурга. Указанный район выбран как самый большой район города по численности населения (более 573 024 человек), представленный, в основном, зданиями позднего советского и постсоветского периода. В связи с тем, что рынок недвижимости отличается консервативной реакцией на внешние изменения на рынке

(до 6 месяцев) для исследования были выбраны номера Бюллетеня недвижимости с №1531 (3 квартал 2012 года) по №1810 (4 квартал 2017 года) с шагом в один квартал. Рассматривалась многомерная случайная величина $(V_{KC}, V(t_1), \dots, V(t_n))$, где V_{KC} – кадастровая стоимость (в тыс.руб. за 1 кв.м.), $V(t_1), \dots, V(t_n)$ – рыночные цены 1 кв.м. в тыс.руб. по Приморскому району города Санкт-Петербурга, на моменты времени t_1, \dots, t_n нумерация по порядковому номеру квартала, начиная от даты первого наблюдения (4 квартал 2012 года). Надо определить по заданному значению кадастровой стоимости наиболее вероятную траекторию движения рыночной цены 1 квадратного метра и дать прогноз на следующие 10 периодов (2,5 года). Значения $V_{KC}, V(t_1), \dots, V(t_n)$ были прологарифмированы и для них были получены вектор средних, вектор стандартных отклонений, и условная ковариационная матрица

$$COV(\vec{Y}|W = w) = COV - \frac{cov(W\vec{Y})^T \times cov(W, \vec{Y})}{\sigma_W^2}$$

Установка связей для определения коэффициентов ковариаций между массивами рыночных и кадастровых данных происходила через создание ключевого поля посредством объединения полей с адресом и общей площадью объекта.

Табл. 2.1.

Вектор средних значений логарифмов переменных $V_{KC}, V(t_1), \dots, V(t_n)$ и вектор их стандартных отклонений*

	КС 2018	4 кв. 2012	1 кв. 2013	2 кв. 2013	3 кв. 2013	4 кв. 2013	1 кв. 2014	2 кв. 2014	3 кв. 2014	4 кв. 2014	1 кв. 2015	2 кв. 2015	3 кв. 2015	4 кв. 2015	1 кв. 2016	2 кв. 2016	3 кв. 2016	4 кв. 2016	1 кв. 2017	2 кв. 2017	3 кв. 2017	4 кв. 2017
средние	4,630	4,489	4,495	4,499	4,517	4,532	4,551	4,584	4,587	4,620	4,631	4,656	4,652	4,618	4,660	4,620	4,626	4,618	4,610	4,630	4,641	4,639
ст.отклонения	0,120	0,167	0,167	0,171	0,184	0,178	0,173	0,186	0,181	0,185	0,193	0,170	0,179	0,192	0,189	0,181	0,181	0,134	0,179	0,187	0,181	0,182

*Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

Условная ковариационная матрица при фиксированном значении $W = w$ ($V_{KC} = v_{KC}$) получена по формуле (3).

Табл. 2.2.

Условная ковариационная матрица (формула(3))*

	4 кв. 2012	1 кв. 2013	2 кв. 2013	3 кв. 2013	4 кв. 2013	1 кв. 2014	2 кв. 2014	3 кв. 2014	4 кв. 2014	1 кв. 2015	2 кв. 2015	3 кв. 2015	4 кв. 2015	1 кв. 2016	2 кв. 2016	3 кв. 2016	4 кв. 2016	1 кв. 2017	2 кв. 2017	3 кв. 2017	4 кв. 2017
4 кв. 2012	0,0202	0,0042	0,0021	0,0037	0,0063	0,0047	0,0009	0,0039	0,0039	0,0012	0,0003	0,0015	-0,0002	0,0021	0,0032	-0,0039	0,0001	-0,0043	0,0023	0,0017	0,0030
1 кв. 2013	0,0042	0,0199	0,0014	0,0035	0,0094	0,0063	0,0019	0,0056	0,0028	-0,0008	0,0007	0,0013	0,0022	0,0055	0,0036	-0,0005	0,0018	-0,0051	0,0032	0,0027	0,0050
2 кв. 2013	0,0021	0,0014	0,0168	0,0103	0,0102	0,0068	0,0013	0,0047	0,0033	0,0005	0,0017	0,0024	0,0012	0,0067	0,0083	-0,0016	0,0028	-0,0051	0,0024	0,0023	0,0035
3 кв. 2013	0,0037	0,0035	0,0103	0,0228	0,0135	0,0089	0,0060	0,0081	0,0064	0,0020	0,0031	0,0046	0,0035	0,0065	0,0091	0,0000	0,0026	-0,0004	0,0081	0,0057	-0,0004
4 кв. 2013	0,0063	0,0094	0,0102	0,0135	0,0278	0,0151	0,0122	0,0119	0,0115	0,0104	0,0105	0,0109	0,0114	0,0118	0,0134	0,0056	0,0075	0,0031	0,0086	0,0101	0,0091
1 кв. 2014	0,0047	0,0063	0,0068	0,0089	0,0151	0,0220	0,0112	0,0108	0,0105	0,0075	0,0095	0,0087	0,0072	0,0111	0,0132	0,0030	0,0073	0,0003	0,0081	0,0063	0,0088
2 кв. 2014	0,0009	0,0019	0,0013	0,0060	0,0122	0,0112	0,0177	0,0080	0,0061	0,0046	0,0061	0,0070	0,0017	0,0095	0,0106	-0,0028	0,0068	-0,0039	0,0036	0,0037	0,0040
3 кв. 2014	0,0039	0,0056	0,0047	0,0081	0,0119	0,0108	0,0080	0,0227	0,0081	0,0078	0,0077	0,0042	0,0042	0,0065	0,0089	-0,0005	0,0060	-0,0064	0,0059	0,0059	0,0037
4 кв. 2014	0,0039	0,0028	0,0033	0,0064	0,0115	0,0105	0,0061	0,0081	0,0225	0,0057	0,0054	0,0049	0,0048	0,0090	0,0127	-0,0004	0,0073	-0,0055	0,0024	0,0075	0,0028
1 кв. 2015	0,0012	-0,0008	0,0005	0,0020	0,0104	0,0075	0,0046	0,0078	0,0057	0,0180	0,0062	0,0043	0,0012	0,0076	0,0114	-0,0008	0,0043	-0,0080	0,0014	0,0024	0,0007
2 кв. 2015	0,0003	0,0007	0,0017	0,0031	0,0105	0,0095	0,0061	0,0077	0,0054	0,0062	0,0226	0,0102	0,0015	0,0091	0,0121	0,0001	0,0067	-0,0042	0,0027	0,0023	0,0028
3 кв. 2015	0,0015	0,0013	0,0024	0,0046	0,0109	0,0087	0,0070	0,0042	0,0049	0,0043	0,0102	0,0267	0,0135	0,0112	0,0140	0,0051	0,0086	0,0042	0,0036	0,0038	0,0042
4 кв. 2015	-0,0002	0,0022	0,0012	0,0035	0,0114	0,0072	0,0017	0,0042	0,0048	0,0012	0,0015	0,0135	0,0227	0,0087	0,0141	0,0027	0,0076	0,0029	0,0038	0,0014	0,0053
1 кв. 2016	0,0021	0,0055	0,0067	0,0065	0,0118	0,0111	0,0095	0,0065	0,0090	0,0076	0,0091	0,0112	0,0087	0,0283	0,0184	0,0076	0,0105	0,0058	0,0087	-0,0013	0,0042
2 кв. 2016	0,0032	0,0036	0,0083	0,0091	0,0134	0,0132	0,0106	0,0089	0,0127	0,0114	0,0121	0,0140	0,0141	0,0184	0,0283	0,0116	0,0133	0,0118	0,0112	0,0118	0,0078
3 кв. 2016	-0,0039	-0,0005	-0,0016	0,0000	0,0056	0,0030	-0,0028	-0,0005	-0,0004	-0,0008	0,0001	0,0051	0,0027	0,0076	0,0116	0,0133	0,0097	0,0011	0,0024	-0,0012	-0,0021
4 кв. 2016	0,0001	0,0018	0,0028	0,0026	0,0075	0,0073	0,0068	0,0060	0,0073	0,0043	0,0067	0,0086	0,0076	0,0105	0,0133	0,0097	0,0140	0,0082	0,0089	0,0059	0,0040
1 кв. 2017	-0,0043	-0,0051	-0,0051	-0,0004	0,0031	0,0003	-0,0039	-0,0064	-0,0055	-0,0080	-0,0042	0,0042	0,0029	0,0058	0,0118	0,0011	0,0082	0,0123	0,0048	-0,0064	0,0047
2 кв. 2017	0,0023	0,0032	0,0024	0,0081	0,0086	0,0081	0,0036	0,0059	0,0024	0,0014	0,0027	0,0036	0,0038	0,0087	0,0112	0,0024	0,0089	0,0048	0,0236	0,0131	0,0114
3 кв. 2017	0,0017	0,0027	0,0023	0,0057	0,0101	0,0063	0,0037	0,0059	0,0075	0,0024	0,0023	0,0038	0,0014	-0,0013	0,0118	-0,0012	0,0059	-0,0064	0,0131	0,0182	0,0078
4 кв. 2017	0,0030	0,0050	0,0035	-0,0004	0,0091	0,0088	0,0040	0,0037	0,0028	0,0007	0,0028	0,0042	0,0053	0,0042	0,0078	-0,0021	0,0040	0,0047	0,0114	0,0078	0,0194

*Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

При каждом фиксированном значении $V_{KC} = v_{KC}$ последовательное применение формул (1)-(4) дает наиболее вероятную траекторию движения во времени рыночной цены 1 квадратного метра жилой недвижимости в Приморском районе г.Санкт-Петербурга. Такие временные ряды могут быть использованы для построения прогноза движения рыночной цены 1 квадратного метра в будущем периоде.

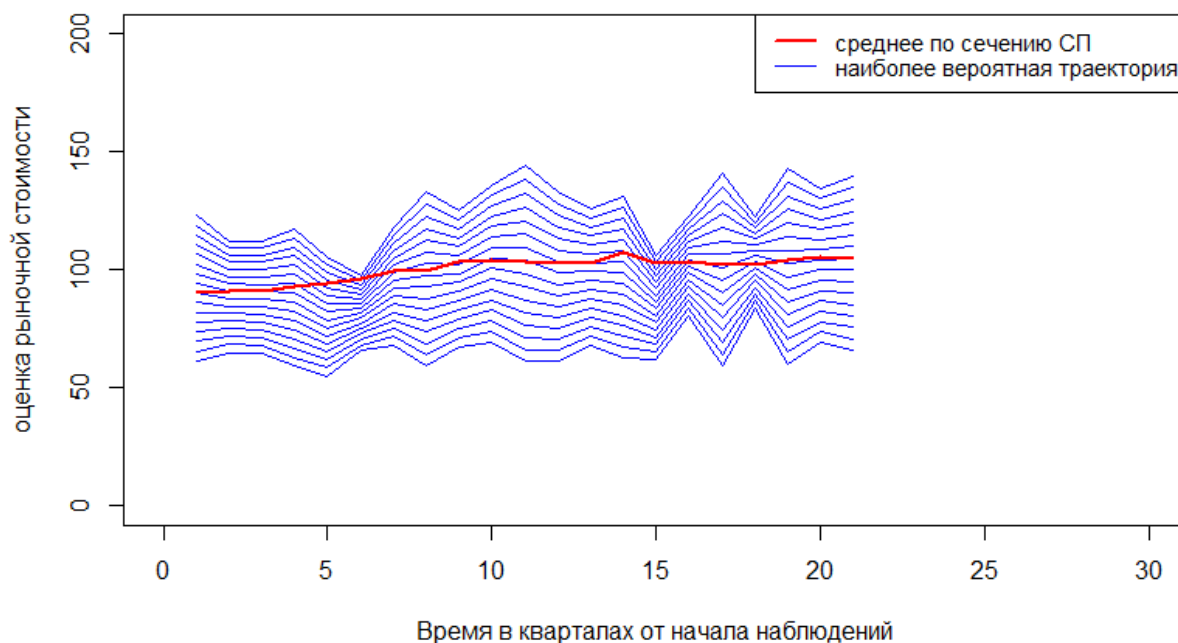


Рис.2.15. Наиболее вероятные траектории для разных значений кадастровой стоимости, установленной в 2018 году

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

Примечание: от 70 тысяч рублей/квадратный метр до 150 тысяч рублей/квадратный метр шаг 5 тысяч рублей/квадратный метр.

На рисунке 2.15. показаны наиболее вероятные линии движения рыночной стоимости 1 квадратного метра в сравнении с общепринятыми оценками (средние значения по рыночным данным) для кадастровых стоимостей от 70 тысяч рублей за 1 квадратный метр до 150 тысяч рублей за 1 квадратный метр с шагом в 5 тысяч рублей (снизу вверх). Полученные временные ряды показывают, что в период 2013-2017 годов. в изучаемом сегменте наблюдалось как однонаправленное движение рынка, например, общий рост во 2 квартале 2016 года (15 квартал от начала наблюдений), так и разнонаправленное движение, например, в 4 квартале 2016 года (17 квартал от начала наблюдений) рыночная стоимость недорогих объектов снизилась, дорогих – повысилась, в 1 квартале 2017 года (18 квартал от начала наблюдений) рыночная стоимость недорогих объектов повысилась, дорогих – снизилась. Следует отметить, что полученные временные ряды, особенно для минимальных и максимальных значений кадастровой стоимости не выглядят как стационарные.

Петроградский район.

Исследованы рыночные данные за период с 4 квартала 2012 по 2 квартал 2017 года по Петроградскому району г. Санкт-Петербурга. Петроградский район города Санкт-Петербурга выбран как самый дорогой район города. 3,4 квартал 2017 года не включен в исследование, прежде всего потому, что в рыночных данных этих периодов присутствует чрезмерно высокая доля дорогой недвижимости, что, по-видимому, отражает изменившуюся структуру района, некоторого застоя на рынке старой недвижимости и повышенного предложения на рынке нового дорогого жилья. Надо определить по заданному значению кадастровой стоимости наиболее вероятную траекторию движения рыночной цены 1 квадратного метра и дать прогноз на следующие периоды. Аналогично, значения $V_{КС}$, $V(t_1), \dots, V(t_n)$ были прологарифмированы и для них были получены вектор средних, вектор стандартных отклонений, и условная ковариационная матрица.

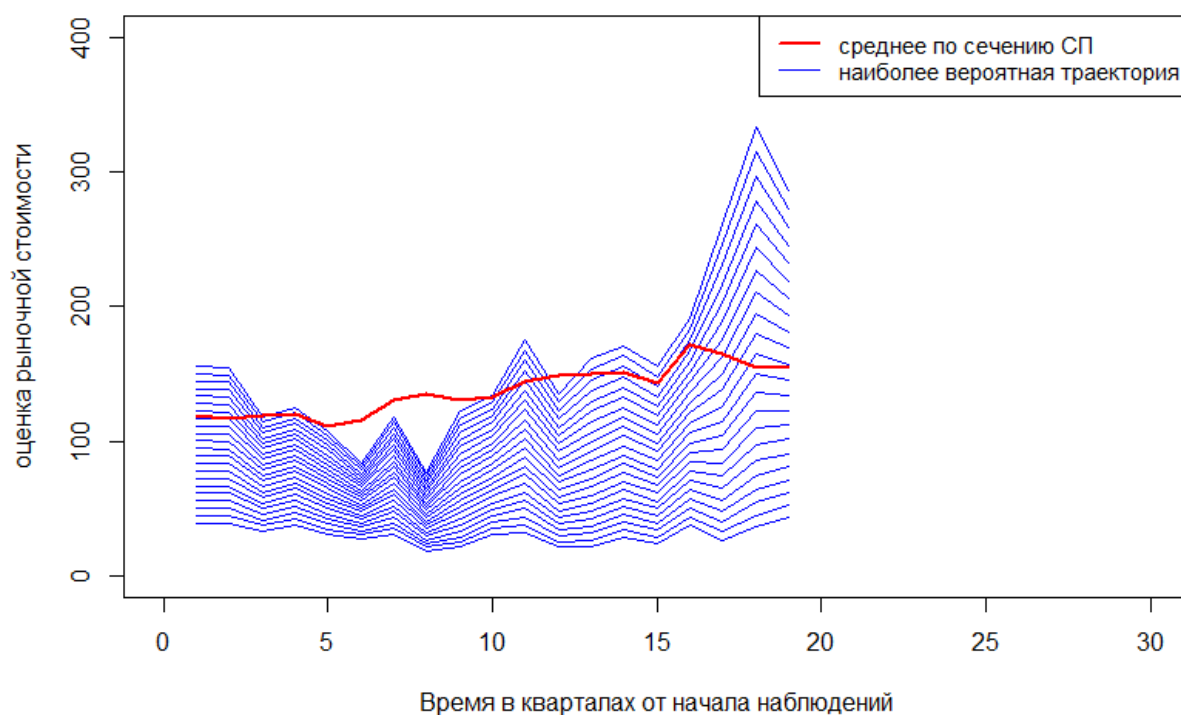


Рис.2.16. Наиболее вероятные траектории для разных значений кадастровой стоимости, установленной в 2018 году

Составлено по: Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 года.

URL: <https://www.bn.ru/> - Сайт Бюллетеня недвижимости.

Примечание: от 60 тысяч рублей за 1 квадратный метр до 280 тысяч рублей за 1 квадратный метр, шаг 10 тысяч рублей за 1 квадратный метр.

На рисунке 2.16. показаны наиболее вероятные линии движения рыночной стоимости 1 квадратного метра в сравнении с общепринятыми оценками (средние значения по рыночным данным) для кадастровых стоимостей от 60 тысяч рублей за 1 квадратный метр до 280 тысяч рублей за 1 квадратный метр с шагом в 10 тысяч рублей (снизу вверх) для Петроградского района. Аналогично Приморскому району в период 2013-2017 годов наблюдается как однонаправленное движение рынка, так и разнонаправленное движение. Все полученные временные ряды для Петроградского района не выглядят как стационарные.

2.2. Анализ временных рядов

Полученные траектории рыночной стоимости во времени представляют собой временные ряды.

Данные по временным рядам могут содержать детерминированные (неслучайные) тренд и сезонность и стохастические (случайные) тренд, сезонность и циклические или остаточные структуры. Детерминированные тренд и сезонность можно оценить, и, таким образом, рассматривать временные ряды, содержащие только остаточный компонент.

Одним из главных свойств для анализа временных рядов является стационарность. Временной ряд считается стационарным, если он не меняет свои характеристики с течением времени. Математически строго стационарным называется временной ряд $\{r_t\}$, если совместное распределение $(r_{t_1}..r_{t_k})$ идентично распределению $(r_{t_1+t}..r_{t_k+t})$ для любых t , где k – некоторое положительное число и $(t_1..t_k)$ – набор от положительных целых k . Другими словами, строгая стационарность требует, чтобы совместное распределение $(r_{t_1+t}..r_{t_k+t})$ оставалось неизменным во времени. Слабо стационарным называется временной ряд $\{r_t\}$, если математическое ожидание $E(r_t)=const$ и $cov(r_t, r_{t-l})=\gamma_l$ и зависит только от l , где l – целое число, т.е это функция, которая зависит только от лага (l). Таким образом, нестрогая стационарность означает, что k значений временного ряда колеблются с постоянным уровнем изменчивости.⁵⁰

В программе RStudio можно проверить временные ряды на стационарность при помощи трёх тестов: Дики-Фуллера, KPSS и PP.

- Тест Дики-Фуллера(`adf.test`)

Нулевая гипотеза состоит в том, что временной ряд является нестационарным. При представлении временного ряда через уравнение с лагами (L), он имеет один или больше единичных корней. Наличие единичного корня также можно считать наличием стохастического тренда. Если результаты теста для данных выдают значение p -value больше критического значения (обычно равного 0,05), это означает нестационарность временного ряда, так как нельзя отвергнуть нулевую гипотезу в пользу альтернативной.

Чтобы избавиться от нестационарности (в данном случае от стохастического тренда), можно продифференцировать временной ряд, то есть взять первую разность. Тогда интегрированный временной ряд $\{r_t\}$ будет иметь вид:

$$\Delta r_t = r_t - r_{t-1}$$

В зависимости от того, сколько единичных корней имеет решение уравнения с лагами, столько раз нужно взять разность, чтобы привести временной ряд к стационарному виду.

- Kwiatkowski-Phillips-Schmidt-Shin тест (`kpss.test`)

Нулевая гипотеза заключается в том, что исходный ряд тренд стационарен ($null = \text{“Trend”}$). Тренд стационарный временной ряд не является слабо стационарным, и, чтобы

⁵⁰ Tsay, R. S. (2005). Analysis of financial time series (Vol. 543). John Wiley & Sons.

привести его к такому виду, необходимо очистить ряд от тренда. Также стоит отметить, что нельзя дифференцировать тренд стационарный временной ряд, и, наоборот, нельзя очищать ряд от детерминированного тренда, если он имеет единичные корни, то есть содержит стохастический тренд.

- Phillips—Perron test(pp.test)

Нулевая гипотеза заключается в том, что временной ряд является нестационарным, то есть имеет один или несколько единичных корней.

Для построения линейных моделей также важны автокорреляционная (ACF) и частная автокорреляционная функции (PACF).

Рассмотрим слабо стационарный временной ряд r_t . Если зависимость между переменной r_t и её прошлыми значениями r_{t-i} является значимой, то понятие корреляции сводится к автокорреляции. Тогда коэффициент корреляции между r_t и r_{t-l} называется лаг-автокорреляцией r_t и определяется через ρ_l , которая при условии слабой стационарности является функцией только от l (лага)⁵¹ и имеет вид:

$$\rho_l = \frac{Cov(r_t, r_{t-l})}{Var(r_t)}$$

Таким образом, автокорреляционная функция представляет собой зависимость взаимосвязи между наблюдением и его прошлыми значениями от величины лага.

Частная автокорреляционная функция (PACF) представляет собой коэффициент корреляции между r_t и r_{t-l} , исключая влияние промежуточных значений $r_{t-1}, r_{t-2}, \dots, r_{t-(l-1)}$. PACF является полезной функцией для определения порядка авторегрессии в модели.

AR-модели для данных

Авторегрессионная модель представляет собой линейную модель временного ряда, где наблюдаемое значение зависит от предыдущих значений во времени. В общем виде AR-модель порядка p можно представить, как

$$r_t = \varphi_0 + \varphi_1 r_{t-1} + \dots + \varphi_p r_{t-p} + \alpha_t$$

где p – неотрицательное целое число, φ - коэффициенты модели, а $\{\alpha_t\}$ предполагается как белый шум со средним равным нулю и дисперсией равной δ_α^2 .⁵² На практике порядок p определяется с помощью частной автокорреляционной функции PACF.

$$AR(1): r_t = \varphi_{0,1} + \varphi_{1,1} r_{t-1} + \varepsilon_{1t}$$

$$AR(2): r_t = \varphi_{0,2} + \varphi_{1,2} r_{t-1} + \varphi_{2,2} r_{t-2} + \varepsilon_{2t}$$

$$AR(3): r_t = \varphi_{0,3} + \varphi_{1,3} r_{t-1} + \varphi_{2,3} r_{t-2} + \varphi_{3,3} r_{t-3} + \varepsilon_{3t}$$

$$AR(4): r_t = \varphi_{0,4} + \varphi_{1,4} r_{t-1} + \varphi_{2,4} r_{t-2} + \varphi_{3,4} r_{t-3} + \varphi_{4,4} r_{t-4} + \varepsilon_{4t} \text{ и т.д.}$$

⁵¹ Tsay, R. S. (2005). Analysis of financial time series (Vol. 543). John Wiley & Sons.

⁵² Brooks, C. (2014). Introductory econometrics for finance. Cambridge university press

где $\varphi_{0,j}$ – постоянные переменные, $\varphi_{i,j}$ – коэффициенты r_{t-i} и $\{\varepsilon_{jt}\}$ – случайные ошибки модели AR(j). Оцененные с помощью метода наименьших квадратов коэффициенты $\hat{\varphi}_{1,1}$, $\hat{\varphi}_{2,2}$, $\hat{\varphi}_{3,3}$ называются соответственно лаг-1, лаг-2, лаг-3 выборочными PACF для r_t . Выборочные PACF для AR-процесса порядка p обладают следующими свойствами:

- $\hat{\varphi}_{p,p}$ приближается к φ_p по мере приближения размера выборки T к бесконечности
- $\hat{\varphi}_{l,l}$ приближается к 0 для всех $l > p$
- Асимптотическая дисперсия $\hat{\varphi}_{l,l}$ равна $1/T$ для $l > p$

Таким образом, выборочная PACF прерывается на лаге p для ряда AR порядка p .⁵³

Чтобы более четко определить порядок p AR-модели используются информационные критерии. Наиболее известный критерий AIC (информационный критерий Акаике). AIC определяется как:

$$AIC = \frac{-2}{T} \ln(\text{likelihood}) + \frac{2}{T} * (\text{количество параметров}),$$

где likelihood – функция правдоподобия, оцененная с помощью метода максимального правдоподобия (суть которого заключается в поиске максимума функции), а T – размер выборки. Второе слагаемое критерия «штрафует» за слишком большое количество параметров в модели.[11.37] Следовательно, из двух моделей с одинаковыми по значению критериями AIC стоит выбирать ту, где порядок p меньше.

Таким образом, чем более правдоподобна подбираемая модель, тем меньше критерий AIC в силу отрицательного знака перед слагаемым, включающим в себя функцию правдоподобия, в формуле критерия. Критерий AIC можно выразить с помощью другой формулы:

$$AIC = \ln(\hat{\sigma}^2) + \frac{2}{T} * (\text{количество параметров}),$$

где $\hat{\sigma}^2$ – дисперсия остатков, которая эквивалентна сумме квадратов остатков, деленная на количество наблюдений T . Она оценивается с помощью метода наименьших квадратов, следовательно, чем более правдоподобна модель, тем меньше первое слагаемое в формуле критерия. Так, независимо от формулы, модель следует выбирать, исходя из минимального значения критерия AIC.

RStudio также выводит результаты по еще двум информационным критериям. Это Байесовский информационный критерий Шварца (BIC/SC) и скорректированный критерий Акаике (AIC_c). Математически их можно выразить через следующие формулы:

$$BIC = \ln(\hat{\sigma}^2) + \frac{\ln T}{T} * (\text{количество параметров})$$

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

⁵³ Brooks, C. (2014). Introductory econometrics for finance. Cambridge university press

Критерий BIC больше «штрафует» за увеличение количества параметров в модели, так как $\ln T \cdot \text{количество параметров}$ возрастает быстрее чем $2 \cdot \text{количество параметров}$. Скорректированный критерий AIC обычно используется для выборок малого размера. Принцип выбора модели по этим критериям такой же, как и по критерию AIC (минимальное значение). «В целом, нельзя сказать, что один критерий превосходит другой».⁵⁴

ARIMA-модели для данных

«Чистый» AR-процесс имеет геометрически убывающую ACF».⁵⁵ Оценивая таким образом коррелограммы ACF для данных, мы не можем утверждать, что именно AR-модели лучше всего описывают их.

Модель ARIMA эквивалентна модели ARMA, то есть используется характеристическое уравнение модели ARMA только на временном ряду, который приведен к стационарному виду. ARIMA порядка (p,d,q) может быть математически записана как:

$$\Delta^d r_t = \varphi_0 + \sum_{i=1}^p \varphi_i r_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t,$$

где d – порядок интеграции модели (количество взятых разностей, чтобы привести данные к стационарному виду), φ_0 - константа, φ_i и θ_i - коэффициенты при AR и MA процессах, $\{\varepsilon_t\}$ -белый шум.

В RStudio можно подобрать порядок модели ARIMA вручную на основе наименьшего значения критерия AIC/BIC или с помощью функции `auto.arima`, которая подбирает порядок автоматически.

Чтобы оценить точность моделей и выбрать наиболее подходящую, необходимо оценить среднеквадратическую ошибку (RMSE – root-mean-squared error), это одна из мер оценки отличий фактических значений от значений, предсказанных моделью.

RMSE параметр является квадратным корнем его средней квадратичной ошибки и вычисляется по формуле:

$$\text{RMSE}(\hat{\theta}) = \sqrt{E((\hat{\theta} - \theta)^2)}, \text{ где } \hat{\theta} - \text{оцененная функция}$$

Чем меньше показатель RMSE, тем меньше отклонение фактического значения от выбранного с помощью модели и тем точнее модель.

⁵⁴ Brooks, C. (2014). Introductory econometrics for finance. Cambridge university press

⁵⁵ Хисамутдинов И.А. Основы экономики и теории рынка. – М.: Магистр, 2010. – 686 с.

ГЛАВА 3. ПРОГНОЗИРОВАНИЕ РЫНОЧНОЙ СТОИМОСТИ

3.1. Прогнозы на основе моделей временных рядов

На основе полученных в ходе предшествующего анализа временных рядов мы можем построить прогнозы для данных при разной соответствующей им кадастровой стоимости, применяя модели для анализа временных рядов.

Приморский район. Временной ряд при кадастровой стоимости, установленной в размере 70 тысяч рублей за 1 квадратный метр

При проверке тестом Дики-Фуллера ряд становится стационарным после взятия второй разности. На рисунке 3.1. представлена коррелограмма для автокорреляционной и частной автокорреляционной функций для второй разности временного ряда, которая поможет определиться в выборе порядка рид.

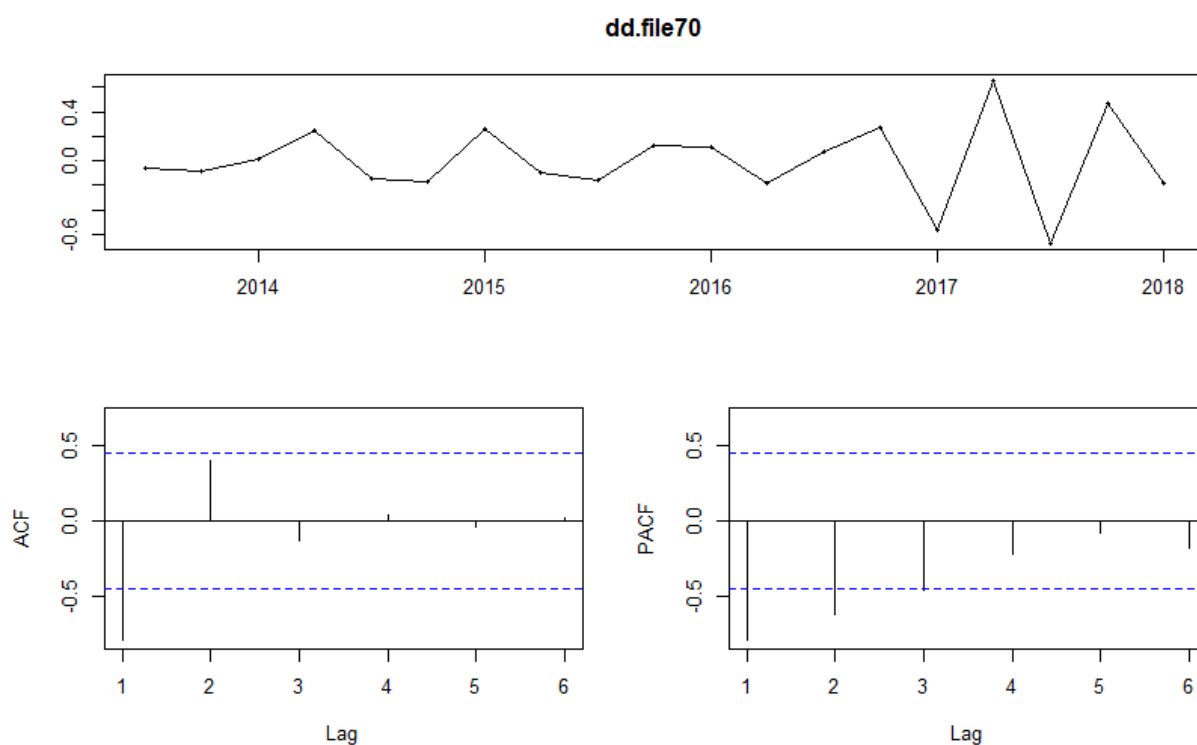


Рис.3.1. Коррелограмма для второй разности временного ряда рыночных цен при кадастровой стоимости, установленной в размере 70 тысяч рублей за 1 квадратный метр

Составлено по: [URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости.

Примечание: **Lag** – количество лагов

ACF - автокорреляционная функция

PACF - частная автокорреляционная функция

Функция `auto.arima` рекомендует модель (3,0,0) для вторых разностей. Среди всех моделей с тремя коэффициентами наименьшим значением информационного критерия

Акаике, Байесовского информационного критерия и наименьшими ошибками RMSE обладает модель (2,2,1) для логарифмов исходных данных.

```
> summary(a2)
Series: lnfile70
ARIMA(2,2,1)

Coefficients:
      ar1      ar2      ma1
    -1.0147  -0.4245  -1.0000
s.e.    0.1988   0.1965   0.1618

sigma^2 estimated as 0.01174:  log likelihood=14
AIC=-19.99  AICc=-17.13  BIC=-16.21

Training set error measures:
              ME      RMSE      MAE      MPE
Training set 0.003376162 0.09458654 0.0753478 0.04216281
              MAPE      MASE      ACF1
Training set 1.79404 0.7592929 -0.1348978
```

Рисунок 3.2. Коэффициенты и статистика для модели ARIMA(2,2,1)

Составлено по: [URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости.

Результат прогноза в логарифмах представлен на рисунке 3.3.

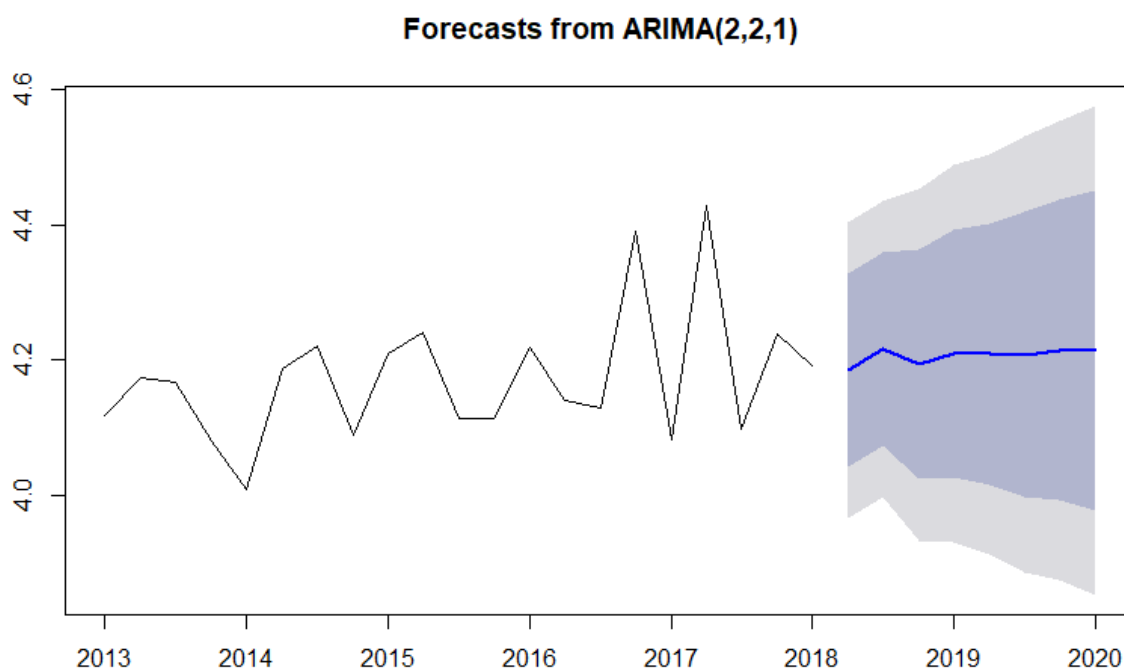


Рис.3.3. Модель ARIMA(2,2,1) и прогноз в логарифмах для временного ряда рыночных цен при кадастровой стоимости, установленной в размере 70 тысяч рублей за 1 квадратный метр

Примечание: горизонтальная ось — время(годы), вертикальная ось — значение цены в логарифмах.

Также необходимо проверить остатки на нормальность для того, чтобы выполнить одно из требований к использованию такого рода модели. Для этой задачи использовался тест Шапиро-Уилка.


```
> shapiro.test(residuals(a2))

Shapiro-Wilk normality test

data:  residuals(a2)
W = 0.93134, p-value = 0.1464
```

Рис. 3.4. Результаты теста Шапиро-Уилка для ARIMA(2,2,1)

Исходя из результатов теста можно заключить, что вероятность совершить ошибку первого рода, то есть отвергнуть верную нулевую гипотезу достаточно высока, выше установленного уровня значимости $= 0,05$, поэтому у нас нет оснований отвергнуть нулевую гипотезу о нормальном распределении остатков.

На рисунке 3.5. представлен результат преобразования прогноза для логарифмов в прогноз по ценам, с учетом того, что в условиях модели логарифмически нормального распределения, каждое сечение прогноза в ценах следует рассматривать как логарифмически нормальное (или нормальное в логарифмах). Это означает, что если прогноз для логарифмов на фиксированный момент времени распределен нормально со средним μ и стандартным отклонением σ , то при построении прогноза в ценах, следует учитывать, что для цен наиболее вероятное значение (мода) равно $e^{\mu-\sigma^2}$, медианное – равно e^{μ} , значение, соответствующее оценке математического ожидания равно $e^{\mu+\frac{1}{2}\sigma^2}$.

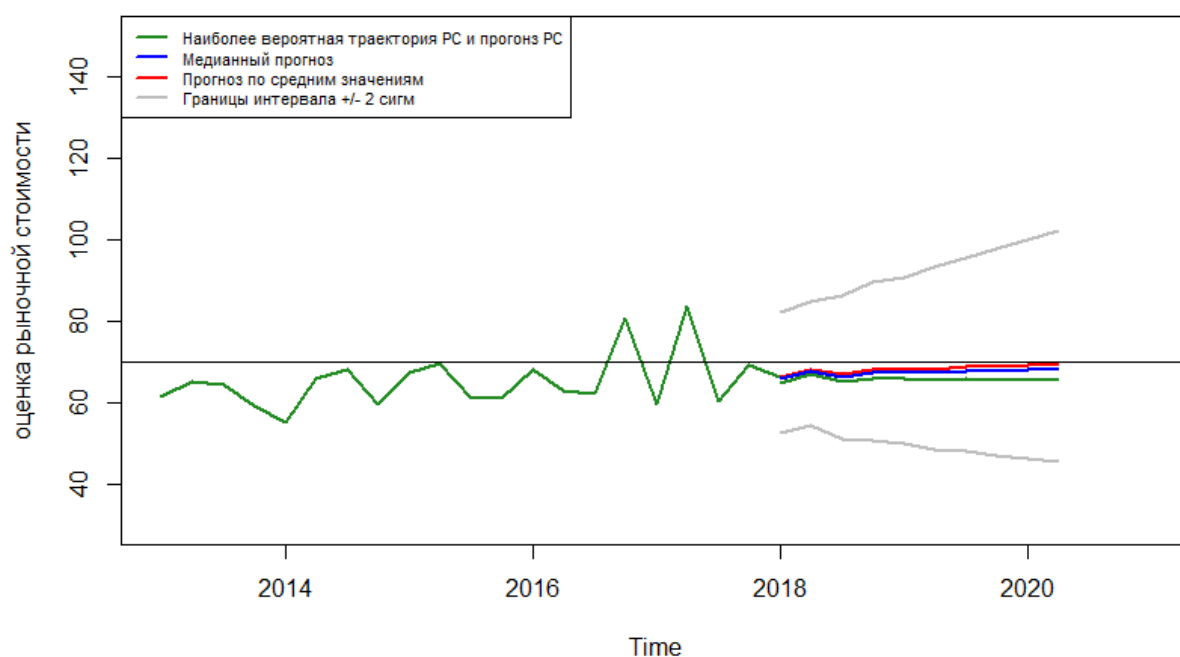


Рис.3.5. Модель ARIMA(2,2,1) и прогноз в реальных ценах для временного ряда рыночных цен при кадастровой стоимости, установленной в размере 70 тысяч рублей за 1 квадратный метр

При переводе прогноза из логарифмов в реальные цены следует учитывать, что рыночная стоимость, по определению является наиболее вероятной ценой, т.е. модой соответствующего сечения случайного процесса. Зеленая линия на рисунке показывает прогнозное значение для рыночной стоимости, оцениваемой по моде. Голубая линия показывает прогнозное значение, оцененное по медиане, а красная – по математическому ожиданию сечений случайного процесса.

Так как, практикующие оценщики часто используют средние арифметические, реже средние геометрические значения, то на рисунке показаны прогнозы по медианным и средним значениям. Но наиболее вероятные значения, отвечающие концепции рыночной стоимости, заложенной в зарубежные стандарты и в ФЗ-135 соответствуют модальному прогнозу. Откуда видно, что следует ожидать снижение рыночной стоимости таких объектов в 2020 году.

Приморский район. Временной ряд при кадастровой стоимости, установленной в размере 100 тысяч рублей за 1 квадратный метр

Рассмотрим поведение временного ряда рыночной стоимости в Приморском районе при более высокой кадастровой стоимости равной 100 тысяч рублей.

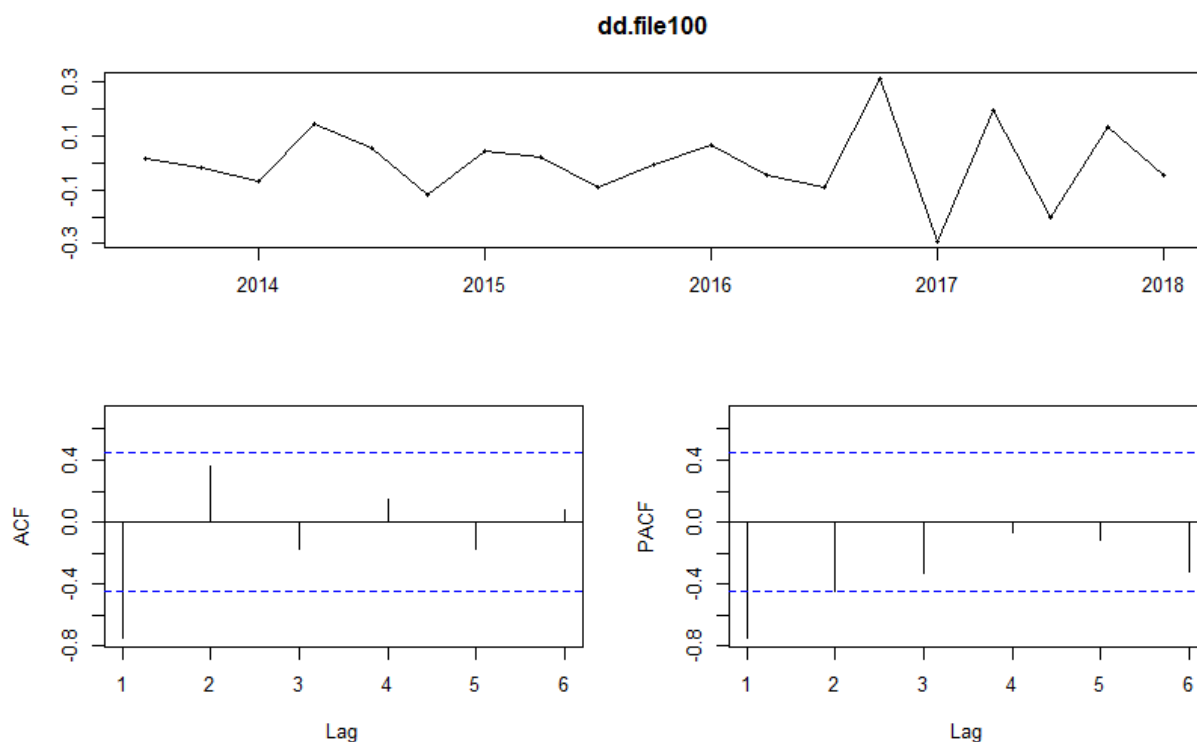


Рис. 3.6. Коррелограмма для второй разности временного ряда рыночных цен при кадастровой стоимости, установленной в размере 100 тысяч рублей за 1 квадратный метр

Составлено по: [URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости.

Примечание: **Lag** – количество лагов

ACF - автокорреляционная функция

PACF - частная автокорреляционная функция

Функция `auto.arima` рекомендует модель (2,0,0) для вторых разностей. Среди всех моделей с двумя коэффициентами наименьшим значением информационного критерия Акаике, Байесовского информационного критерия и наименьшими ошибками RMSE обладает модель (1,2,1) для логарифмов исходных данных.

```
> summary(a2)
Series: lnfile70
ARIMA(1,2,1)

Coefficients:
      ar1      ma1
    -0.4844  -0.9999
s.e.   0.1898   0.1593

sigma^2 estimated as 0.004859: log likelihood=22.7
AIC=-39.4   AICc=-37.8   BIC=-36.56

Training set error measures:
      ME      RMSE      MAE      MPE
Training set 0.007839992 0.06271757 0.04326992 0.1647647
      MAPE      MASE      ACF1
Training set 0.9640597 0.5243794 -0.0863821
```

Рис. 3.7. Коэффициенты и статистика для модели ARIMA(1,2,1)

Результат прогноза в логарифмах представлен на рисунке 3.8.

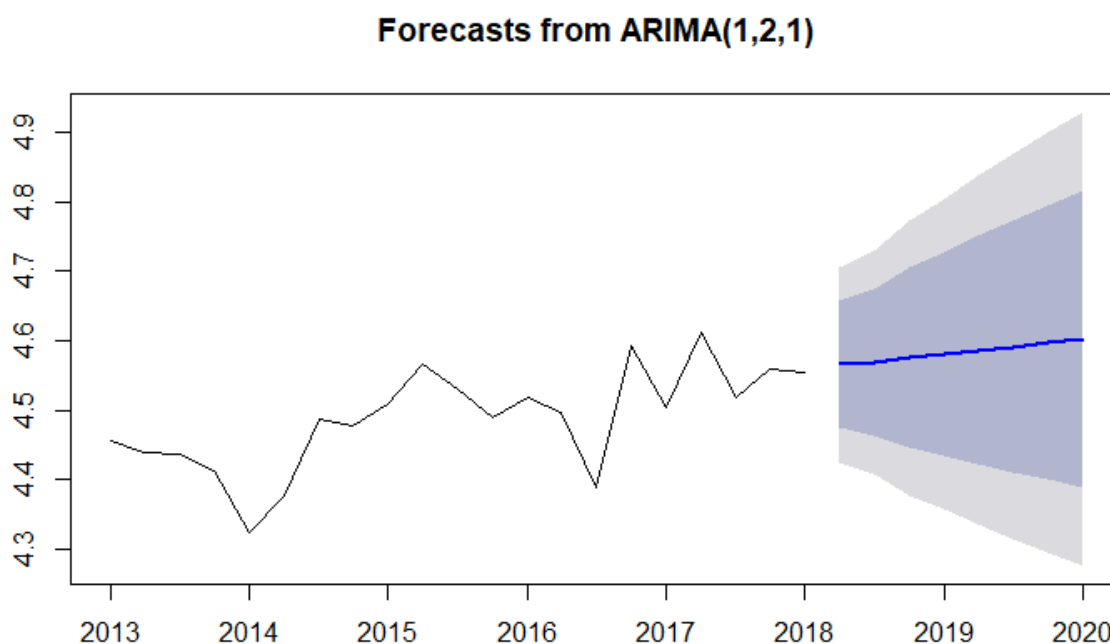


Рис.3.8. Модель ARIMA(1,2,1) и прогноз в логарифмах для временного ряда рыночных цен при кадастровой стоимости, установленной в размере 100 тысяч рублей за 1 квадратный метр

Примечание: горизонтальная ось – время(годы), вертикальная ось – значение цены в логарифмах.

Также необходимо проверить остатки на нормальность для того, чтобы выполнить одно из требований к использованию такого рода модели. Для этой задачи использовался тест Шапиро-Уилка.

```
> shapiro.test(residuals(a2))

Shapiro-Wilk normality test

data:  residuals(a2)
W = 0.93134, p-value = 0.1464
```

Рис.3.9. Результаты теста Шапиро-Уилка для ARIMA(1,2,1)

Исходя из результатов теста можно заключить, что вероятность совершить ошибку первого рода, то есть отвергнуть верную нулевую гипотезу достаточно высока (0,15), выше установленного уровня значимости = 0,05, поэтому у нас нет оснований отвергнуть нулевую гипотезу о нормальном распределении остатков. Условие о нормальности остатков соблюдено.

На *рисунке 3.10.* представлен результат преобразования прогноза для логарифмов в прогноз по ценам, с учетом того, что в условиях модели логарифмически нормального распределения, каждое сечение прогноза в ценах следует рассматривать как логарифмически нормальное (или нормальное в логарифмах).

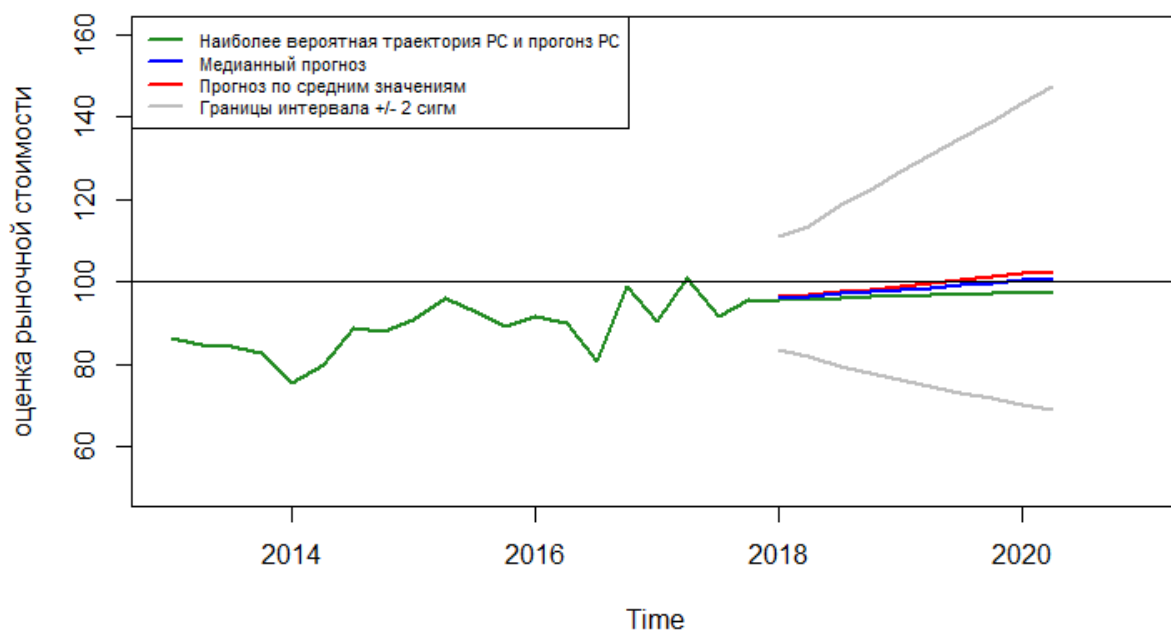


Рис.3.10. Модель ARIMA(1,2,1) и прогноз в реальных ценах для временного ряда рыночных цен при кадастровой стоимости, установленной в размере 100 тысяч рублей за 1 квадратный метр

На рисунке также показаны прогнозы по медианным и средним значениям. Но наиболее вероятные значения, отвечающие концепции рыночной стоимости, заложенной в зарубежные стандарты и в ФЗ-135 соответствуют модальному прогнозу. В этом случае модальный прогноз ведёт себя схоже с медианным прогнозом и прогнозом по среднему значению, но показывает более плавный и сглаженный рост.

В таблице 3 представлены модальные, медианные и средние прогнозные оценки рыночной стоимости для Приморского района Санкт-Петербурга в 3 квартале 2020 года для объектов с кадастровой стоимостью 70,100,150 тысяч рублей за 1 квадратный метр.

Табл. 3.1.

Оценки рыночной стоимости для Приморского района Санкт-Петербурга в 3 квартале 2020 года для объектов с кадастровой стоимостью 70,100,150 тысяч рублей за 1 квадратный метр

Кадастровая стоимость (тыс.руб.за 1 кв.м.)	Наиболее вероятное значение - РС (тыс.руб.за 1 кв.м.)	Медианное значение (тыс.руб.за 1 кв.м.)	Среднее значение (тыс.руб.за 1 кв.м.)
70	64,089	69,912	73,020
100	97,302	100,884	102,724
150	134,847	154,847	165,933

Из таблицы мы можем увидеть, что прогноз по модальным значениям практически соответствует кадастровой стоимости в 3 квартале 2020 года. С увеличением кадастровой стоимости модальный прогноз в отличие от медианного и среднего значения начинает всё сильнее снижаться, что подтверждает предположение о том, что стоит ожидать уменьшения стоимости таких объектов.

Далее попробуем применить выбранную модель к другому району города, структура и специфика которого сильно отличаются от Приморского района.

Петроградский район. Временной ряд при кадастровой стоимости, установленной в размере 150 тысяч рублей за 1 квадратный метр

Исследованы рыночные данные за период с 4 квартала 2012 по 2 квартал 2017 года по Петроградскому району города Санкт-Петербурга. Петроградский район города Санкт-Петербурга выбран как самый дорогой район города. 3,4 квартал 2017 года не включен в исследование, прежде всего потому, что в рыночных данных этих периодов присутствует чрезмерно высокая доля дорогой недвижимости, что, по-видимому, отражает изменившуюся структуру района, некоторого застоя на рынке старой недвижимости и повышенного предложения на рынке нового дорогого жилья.

При проверке тестом Дики-Фуллера ряд становится стационарным после взятия второй разности. Библиотечная функция `auto.arima` рекомендует порядок $p=1$, $q=1$. На рисунке 16 представлена коррелограмма для автокорреляционной и частной автокорреляционной функций для второй разности временного ряда, которая поможет определиться в выборе порядка p и q .

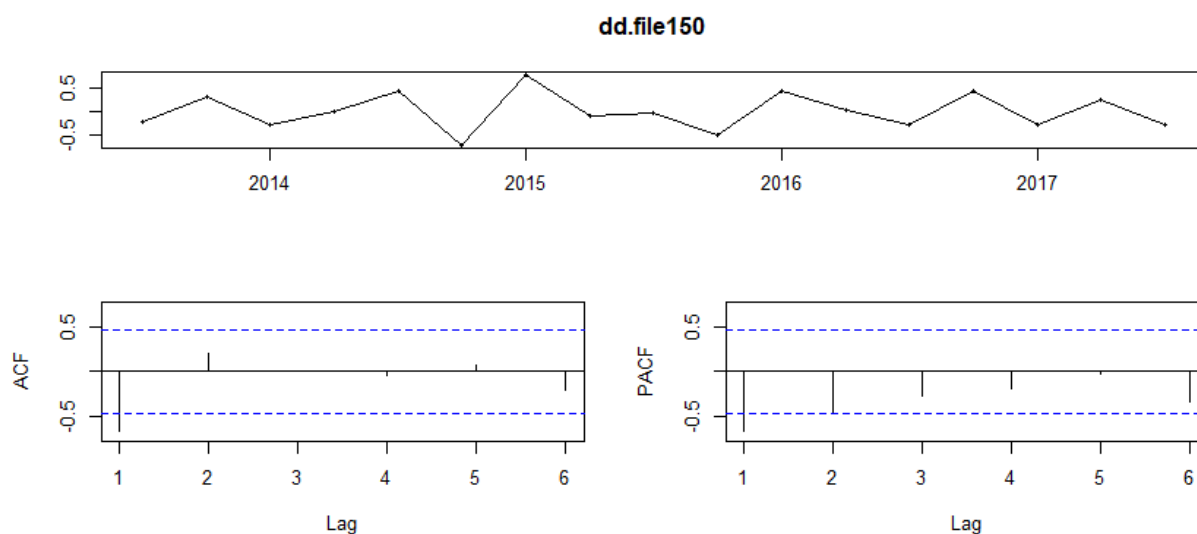


Рис.3.11. Коррелограмма для второй разности временного ряда рыночных цен при кадастровой стоимости, установленной в размере 150 тысяч рублей за 1 квадратный метр

Составлено по: [URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости.

Примечание: **Lag** – количество лагов

ACF - автокорреляционная функция

PACF - частная автокорреляционная функция

Среди всех протестированных моделей наименьшим значением информационного критерия Акаике, Байесовского информационного критерия и наименьшими ошибками RMSE обладает модель (1,2,1) для логарифмов исходных данных.

```
ARIMA(1,2,1)

Coefficients:
      ar1      mal
    -0.4447  -0.7723
s.e.    0.2241   0.1696

sigma^2 estimated as 0.05508:  log likelihood=0.72
AIC=4.55  AICc=6.4  BIC=7.05

Training set error measures:
              ME      RMSE      MAE
Training set 0.05347328 0.2085222 0.169708
              MPE      MAPE      MASE
Training set 1.125143 4.037263 0.5262675
              ACF1
Training set -0.1995718
```

Рис.3.12. Коэффициенты и статистика для модели ARIMA(1,2,1)

Результат прогноза в логарифмах представлен на рисунке 3.13.

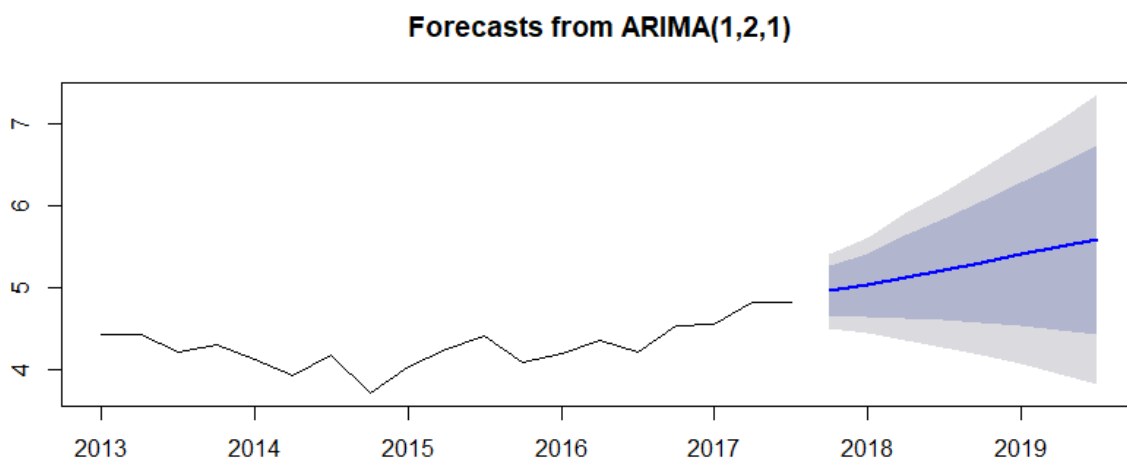


Рис.3.13. Временной ряд, составленный для наиболее вероятной траектории логарифмов цен за период с 4 квартала 2012 г. по 4 квартал 2017 года, в Петроградском районе Санкт-Петербурга и прогноз до 3 квартала 2019 года при кадастровой стоимости равной 150 тысяч рублей за 1 квадратный метр

Примечание: горизонтальная ось – время(годы), вертикальная ось – значение цены в

логарифмах.

Ниже представлены результаты теста Шапиро-Уилка для проверки соответствия распределения остатков в модели нормальному.

```
> shapiro.test(residuals(a2))

Shapiro-Wilk normality test

data: residuals(a2)
W = 0.9761, p-value = 0.8882
```

Рис.3.14. Результаты теста Шапиро-Уилка для ARIMA(1,2,1)

Исходя из результатов теста можно заключить, что вероятность совершить ошибку первого рода, то есть отвергнуть верную нулевую гипотезу высока (0,8882), намного выше установленного уровня значимости = 0,05 и превышает 80 %, поэтому у нас нет оснований отвергнуть нулевую гипотезу о нормальном распределении остатков.

На рисунке 3.15. представлен результат преобразования прогноза для логарифмов в прогноз по ценам, с учетом особенностей модели совместного логарифмически нормального распределения.

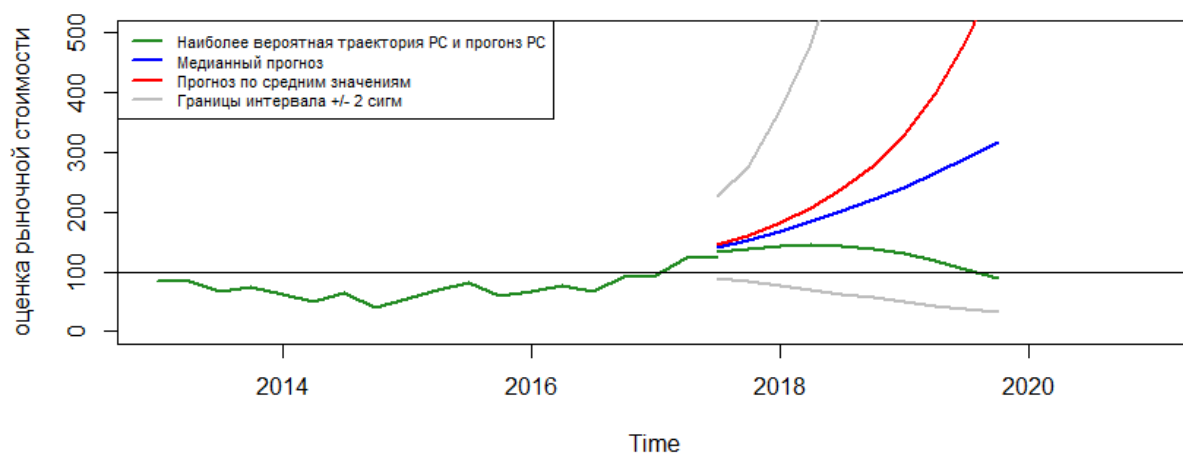


Рис.3.15. Модель ARIMA(1,2,1) и прогноз в реальных ценах для временного ряда рыночных цен при кадастровой стоимости, установленной в размере 150 тысяч рублей за 1 квадратный метр

Исходя из полученных результатов, не представляются реалистичными прогнозы по медианным значениям или по математическим ожиданиям. Временные ряды, полученные для Петроградского района, отличаются высокой волатильностью, модель ARIMA (1,2,1) не

имеет «длинной» памяти, ошибки прогноза быстро нарастают. Однако, участники рынка ориентируются на средние значения, значит склонны надеяться на повышение. Но при увеличивающейся волатильности может оказаться, что средние и медианные оценочные значения растут, а модальные сначала растут, а потом падают, как и показано на рисунке 6. Подобный эффект, возможно, указывает на формирующийся на рынке ценовой «пузырь». Возможно, в этом районе города следует ожидать коррекции рынка вниз. В то же время видно, что предсказанная модальная оценка рыночной стоимости в начале 2018 года, практически соответствует кадастровой стоимости, которая определялась как рыночная на дату 01.01.2018 г. Дальнейшая коррекция рыночной стоимости для объектов с фиксированной кадастровой стоимостью 150 тысяч рублей за 1 квадратный метр ожидается в сторону снижения.

Табл. 4.

Модальные оценки рыночной стоимости для объектов в Петроградском районе Санкт-Петербурга с 3 квартала 2017 года по 4 квартал 2019 года с кадастровой стоимостью 150 тысяч рублей за 1 квадратный метр

3 кв. 2017 г.	4 кв. 2017 г.	1 кв. 2018 г.	2 кв. 2018 г.	3 кв. 2018 г.	4 кв. 2018 г.	1 кв. 2019 г.	2 кв. 2019 г.	3 кв. 2019 г.	4 кв. 2019 г.
133,990	139,364	143,953	145,403	143,730	138,553	129,893	118,127	103,910	88,181

Очевидно, что точность прогноза, построенного по модели ARIMA(1,2,1) с течением времени падает, но подобное поведение прогноза (рост, а затем падение) может указывать на формирующийся на данном секторе рынка ценовой «пузырь». В этом районе города на текущий момент времени он может объясняться повышенным предложением новых объектов, проектирование и строительство которых началось намного раньше и предусматривало создание объектов класса «бизнес» и «элита» (типы домов, существующие в кадастровой базе данных). Таким образом, продавцы рассчитывают на высокие цены продаж, в то время как платежеспособный спрос снизился.

Петроградский район. Временной ряд при кадастровой стоимости, установленной в размере 220 тысяч рублей за 1 квадратный метр

Рассмотрим поведение временного ряда рыночной стоимости при более высокой соответствующей ей кадастровой стоимости.

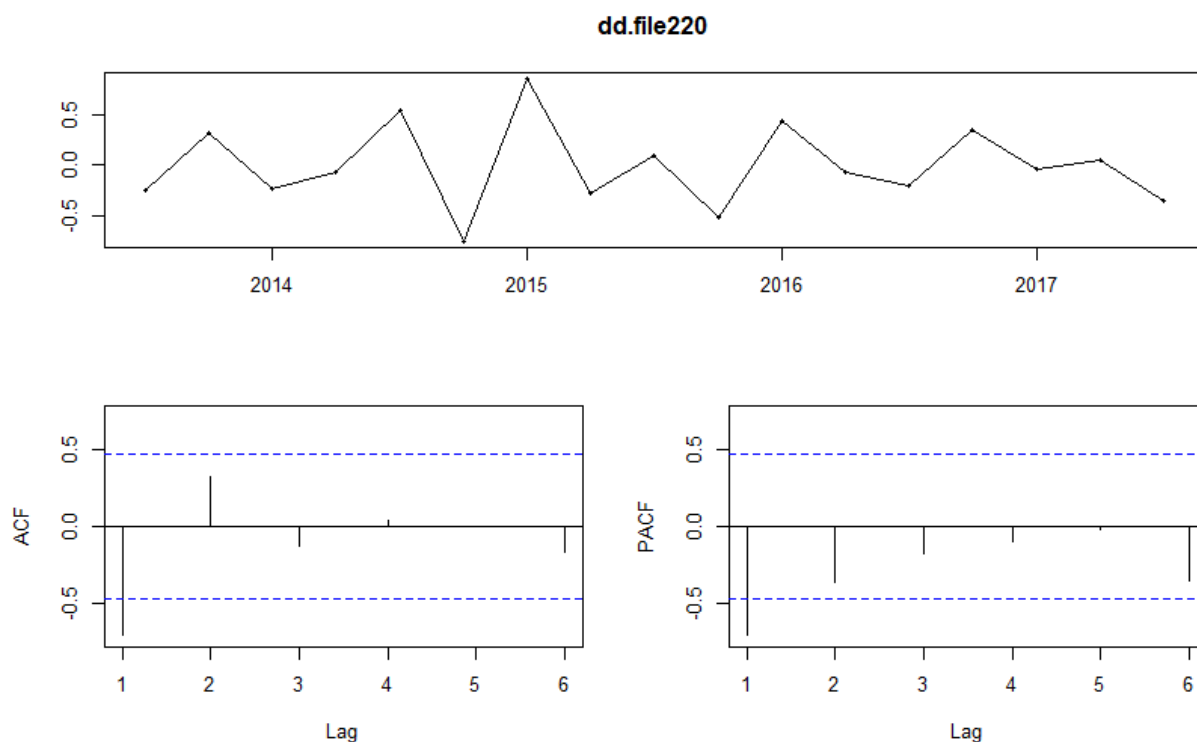


Рис.3.16. Коррелограмма для второй разности временного ряда рыночных цен при кадастровой стоимости, установленной в размере 220 тысяч рублей за 1 квадратный метр

Составлено по: [URL: https://www.bn.ru/](https://www.bn.ru/) - Сайт Бюллетеня недвижимости.

Примечание: **Lag** – количество лагов

ACF - автокорреляционная функция

PACF - частная автокорреляционная функция

Коррелограмма даёт нам основание предполагать, что порядок автокорреляционной и частной автокорреляционной функций не превышает 1. Пользовательская функция `auto.arima` рекомендует модель (1,0,1) для вторых разностей

Среди всех моделей с двумя коэффициентами она обладает наименьшим значением информационного критерия Акаике, Байесовского информационного критерия, наименьшими ошибками RMSE для логарифмов исходных данных.

```

ARIMA(1,2,1)

Coefficients:
          ar1          ma1
      -0.4948   -0.7303
s.e.    0.2270    0.1912

sigma^2 estimated as 0.05716:  log likelihood=0.44
AIC=5.12   AICc=6.97   BIC=7.62

Training set error measures:
              ME      RMSE      MAE      MPE
Training set 0.05389635 0.2124205 0.1782904 1.056587
              MAPE      MASE      ACF1
Training set 3.797983 0.4985576 -0.1687501

```

Рис. 3.17. Коэффициенты и статистика для модели ARIMA(1,2,1)

Результат прогноза в логарифмах представлен на рисунке 3.18.

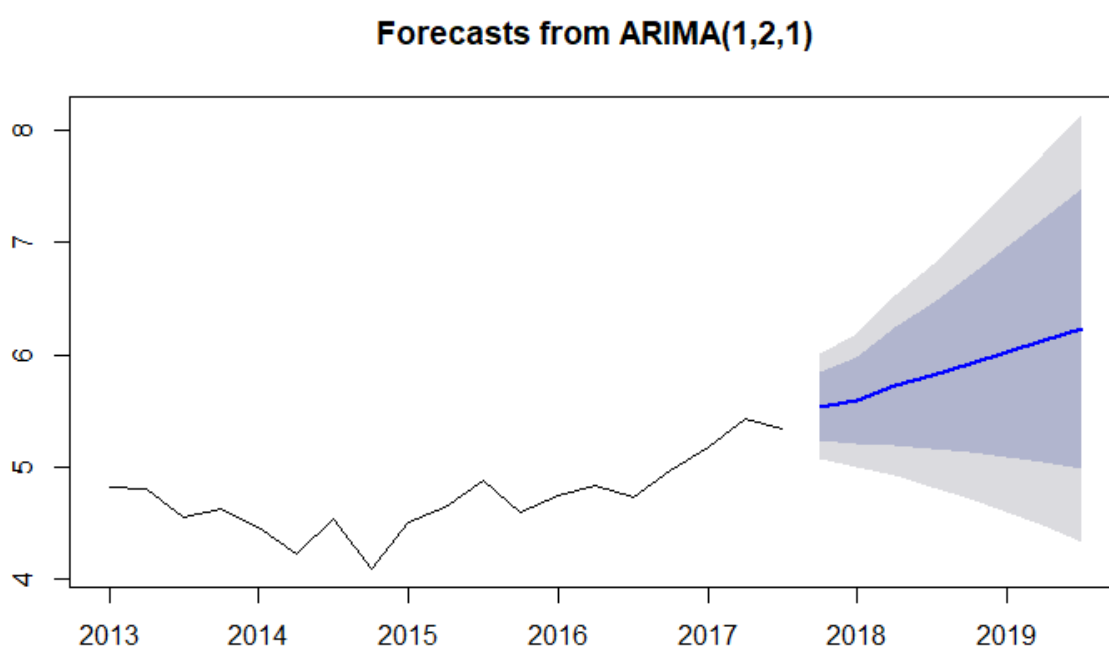


Рис.3.18. Временной ряд, составленный для наиболее вероятной траектории логарифмов цен за период с 4 квартала 2012 г. по 4 квартал 2017 года, в Петроградском районе Санкт-Петербурга и прогноз до 3 квартала 2019 года при кадастровой стоимости равной 220 тысяч рублей за квадратный метр

Примечание: горизонтальная ось — время(годы), вертикальная ось — значение цены в логарифмах.

Ниже представлены результаты теста Шапиро-Уилка для проверки соответствия остатков в модели нормальному распределению.

```
> shapiro.test(residuals(a2))

Shapiro-Wilk normality test

data:  residuals(a2)
W = 0.95255, p-value = 0.4361
```

Рис.3.19. Результаты теста Шапиро-Уилка для ARIMA(1,2,1)

Исходя из результатов теста можно заключить, что вероятность совершить ошибку первого рода, то есть отвергнуть верную нулевую гипотезу высока (0,4361), намного выше установленного уровня значимости $= 0,05$ и достигает почти 50 %, поэтому у нас нет оснований отвергнуть нулевую гипотезу о нормальном распределении остатков.

На рисунке 3.20. представлен результат преобразования прогноза для логарифмов в прогноз по ценам, с учетом особенностей модели совместного логарифмически нормального распределения.

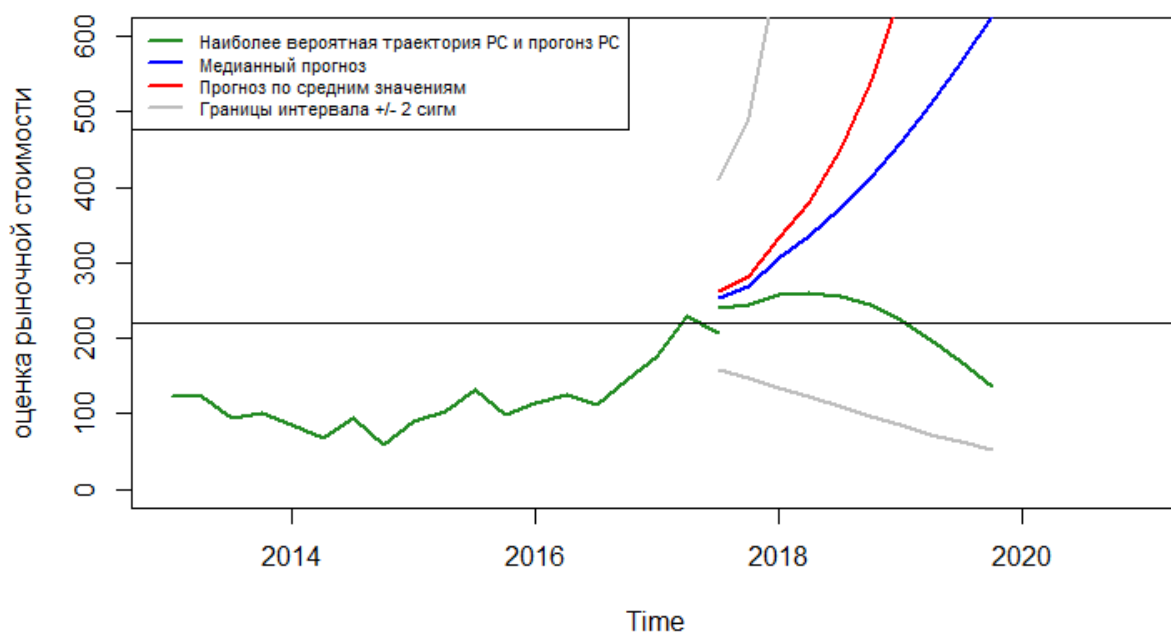


Рис.3.20. Модель ARIMA(1,2,1) и прогноз в реальных ценах для временного ряда рыночных цен при кадастровой стоимости, установленной в размере 220 тысяч рублей за 1 квадратный метр

По результатам применения модели можно заметить, что с увеличением цен замеченные ранее паттерны в поведении временного ряда рыночной стоимости становятся

более выразительными. Прогноз по среднему арифметическому и среднему геометрическому (медиана) и далее показывает свою несостоятельность. Прогноз по модельным значениям цены подтверждает ранее выдвинутую гипотезу о способности модели предсказывать цены-пузыри: линия моды сначала слегка увеличивается, затем начинает достаточно резко падать.

Аналогичные результаты получены и для других значений кадастровой стоимости в Петроградском районе.

ЗАКЛЮЧЕНИЕ

С введением в кадастровые базы кадастровой стоимости, определенной как рыночная на дату оценки по всему массиву объектов недвижимости одновременно, в РФ появилась уникальная возможность мониторинга и прогнозирования рыночной стоимости для любого объекта недвижимости, прошедшего кадастровый учет, даже если такой объект ни разу не выставлялся в рыночных листингах, поскольку такая база представляет по сути срез рынка.

Основной проблемой при обработке рыночных данных и сопоставлении их с данными кадастровых баз была выявлена неструктурированность рыночных данных и отсутствие в них уникального идентификатора – кадастрового номера. Возможным решением проблемы является введение кадастрового номера в рыночные листинги как обязательного параметра любого объявления о продаже, покупке, найме или сдаче в аренду объектов недвижимости. Это позволило бы создавать более быстрые и удобные для пользователей машинные алгоритмы и функции для анализа рынка недвижимости в системах поддержки принятия решений в оценочных компаниях, страховых фирмах, банках и других организациях и учреждениях, нуждающихся в анализе и прогнозировании цен на рынке недвижимости. Кадастровый номер – это уникальное ключевое поле, если выражаться в терминах баз данных, по которому может быть определена не только кадастровая стоимость объекта, но и оценены его рыночная стоимость, ставка дисконтирования, скидка на торг и другие показатели на рынке недвижимости. Другое решение проблемы, менее удобное, оформление адреса и занесение его в листинги по типу адреса в кадастровых базах. Как было выявлено в ходе исследования, такое изменение позволило бы сократить время на предварительную подготовку данных и процесс их анализа.

В ходе теоретической составляющей исследования было изучено достаточно большое количество источников, преимущественно зарубежных авторов. Изучение классической литературы по оценке недвижимости известных авторов-оценщиков и работ, берущих за основу предположение о логарифмически нормальном распределении цен на рынке, сформировало представление о степени изученности вопроса в российском научном сообществе. В зарубежных исследованиях широко применяются современные методы обработки и прогнозирования рыночных данных. Часто авторами используются нетрадиционные, а более инновационные методы, в том числе методы машинного обучения. Работы, берущие за основу индексы цен, разрабатывают свои собственные, противопоставляя их прогнозируемую способность традиционному индексу Кейса-Шиллера.

В процессе подготовки данных было определено к какому типу они относятся, какими методами и каким программным обеспечением лучше всего воспользоваться для их

обработки. В результате была разработана функция в соответствующей VІпрограмме, которая позволила решить проблему отсутствия кадастрового номера в рыночных данных.

В ходе обработки данных была выявлена проблема большого объёма данных, которая привела к невозможности использования традиционных тестов проверки на нормальность распределения. Эта проблема была решена методом последовательного извлечения из выборки цен подвыборок малого объёма (500) и применением теста на проверку нормальности, эта процедура повторялась 1000 раз. При рассмотрении распределения результатов теста на нормальность были приняты гипотезы о логарифмически нормальном распределении стоимости и о совместной нормальности логарифмов цен.

Также в исследовании была использована модель многомерного логарифмически нормального закона распределения ансамбля цен на объекты жилой недвижимости в равноотстоящие моменты времени и кадастровой стоимости. Результатом стал алгоритм оценки наиболее вероятной траектории рыночной стоимости исследуемого объекта на основе многомерного условного логарифмически нормального распределения цен при заданном значении кадастровой стоимости.

В ходе практической части к логарифмам полученных временных рядов применялась хорошо разработанная и изученная модель прогнозирования временных рядов ARIMA, возврат от логарифмированных цен к реальным проводится с учетом особенностей логарифмически нормального распределения. Результаты сравнивались с медианными оценками и оценками, полученными по средним значениям. В результате были получены реалистичные прогнозы рыночной стоимости на примере одного района на разных интервалах соответствующей кадастровой стоимости и выявлены основные тенденции в движении цен на рынке недвижимости. При исследовании идентичной методики на другом районе города, который принципиально отличается по типам жилья и структуре застройки были получены другие результаты. Здесь дальнейшего внимания и исследований заслуживает вопрос о возможности предвидения образования цен-пузырей на основе прогнозирования наиболее вероятных траекторий рыночных стоимостей как условных модальных значений многомерных распределений.

В процессе исследования была достигнута цель работы и выполнены все поставленные задачи. Результат, как было заявлено в начале работы, может быть использован в автоматизированных системах оценочных компаний и далее использоваться другими заинтересованными лицами и структурами на рынке недвижимости. Таким образом, выводы и полученные результаты имеют практическую значимость для бизнес-пользователей и могут быть применены в ходе решения определённых бизнес-задач.

В сфере оценки недвижимости в Российской Федерации ещё достаточно много непроработанных областей, требующих внимания государства, для выработки определённых стандартов и методик работы с ними. Данные по основным числовым характеристикам рынка недвижимости все ещё не в полной мере собраны и структурированы для работы с ними. Это замедляет темпы развития исследований в области недвижимости, в том числе в области прогнозирования динамических характеристик рынка. Тем не менее, такое состояние делает область рынка недвижимости широким полем для исследований и выработки инновационных методов и открытий, которые принесут выгоду не только профессиональным участникам рынка, но и массовым потребителям жилья.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

Нормативно-правовые акты

1. Федеральный закон от 29.07.1998 N 135-ФЗ (ред. от 29.07.2017) «Об оценочной деятельности в Российской Федерации»
2. Федеральный Стандарт Оценки №1 : Приказ Минэкономразвития России "Цель оценки и виды стоимости (ФСО N 1)" от 20 июля 2007 года N 255
3. Федеральный Стандарт Оценки №2 : Приказ Минэкономразвития России "Общие понятия оценки, подходы к оценке и требования к проведению оценки (ФСО N 2)" от 20 июля 2007 года N 256

Книги

4. Грибовский С.В., Математические методы оценки стоимости недвижимого имущества: учеб. пособие / С.В. Грибовский, С.А. Сивец; под ред. С.В. Грибовского, М.А. Федотовой - М.: Финансы и статистика, 2008. - 368 с
5. Грибовский С.В. Оценка стоимости недвижимости: Учебное пособие. — 2-е изд., испр. и доп. — М.: ООО «Про-Аппрайзер» Онлайн, 2016. — 464 с.
6. Е.С. Озеров, С.В. Пупенцова., Управление стоимостью и инвестиционным потенциалом недвижимости - СПб : Изд-во Политехн. ун-та, 2015 . – 600 с.
7. Хисамутдинов И.А. Основы экономики и теории рынка. – М.: Магистр, 2010. – 686 с.
8. Brooks, C. (2014). Introductory econometrics for finance. Cambridge university press
9. Damodaran A. (2012). Investment Valuation: Tools and Techniques for Determining the Value of Any Asset (Vol. 666). John Wiley & Sons. – 992 p.
10. Sayce S., Smith J., Cooper R., Venmore-Rowland P. (2009). Real Estate Appraisal: From Value to Worth. John Wiley & Sons. – 352 p.
11. Tsay, R. S. (2005). Analysis of financial time series (Vol. 543). John Wiley & Sons.

Статьи в журналах

12. Ласкин М.Б. Корректировка рыночной стоимости по ценообразующему фактору «площадь объекта». Имущественные отношения в Российской Федерации. 2017. № 8(191), С.86-99.
13. Ласкин М.Б., Гадасина Л.В. Как определить кадастровую стоимость // Имущественные отношения в Российской Федерации. 2018. № 3. 42-53 с.
14. Ласкин М.Б., Пупенцова С.В. Логарифмическое распределение цен на объекты недвижимости // Имущественные отношения в Российской Федерации, №5 (15)/2014. – с.52-59

15. Ласкин М.Б., Русаков О.В., Джаксумбаева О.И., Ивакина А.А. Особенности формирования величины рыночной стоимости недвижимости при логарифмически нормальном распределении цен // Имущественные отношения в Российской Федерации, №2 (173)/2016. – с.40-50.
16. Лемешко Б. Ю. К вопросу статистического анализа больших данных / Б. Ю. Лемешко, С. Б. Лемешко, М. А. Семёнова // Вестник Томского государственного университета. Управление, вычислительная техника и информатика - 2018. – № 44. – С. 40–49.
17. Русаков О. В. Стохастическая модель ценообразования на рынке недвижимости: формирование логнормальной генеральной совокупности / О. В. Русаков, М. Б. Ласкин, О. И. Джаксумбаева // Вестник УМО. 2015. № 5.
18. Русаков О.В., Ласкин М.Б., Джаксумбаева О.И., Ивакина А.А. Особенности рыночной стоимости на рынке недвижимости при логарифмически нормальном распределении // Имущественные отношения в Российской Федерации. 2016. № 2(173). С. 40–50.
19. Русаков О.В., Ласкин М.Б., Джаксумбаева О.И., Стабровская К.Ю. Определение скидки на торг по статистическим данным. Вестник гражданских инженеров. 2016. №2, С. 268-284.
20. Adams Z., Füss R. Macroeconomic Determinants of International Housing Markets. Journal of Housing Economics 19(1). 2010: pp. 38-50.
21. Antipov E., Pokryshevskaya E. (2010). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. Expert Systems with Applications 39(2). 2012: pp. 1772-1778.
22. Anselin, L., Lozano-Gracia N. Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality, Empirical Economics 34 (1), 2008: pp. 5–34.
23. Barry J.R., Ellisy E.A., Kassaby A., Redfearnz C.L., Srinivasany N., Voris K.B. Home Price Index: A Machine Learning Methodology. International Journal of Semantic Computing 1(1), 2017: pp. 9-22.
24. Bo Huang, Bo Wu, Michael Barry, Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. International Journal of Geographical Information Science 24(3). 2010: pp. 383-401.
25. Bourassa S.C., Hoesli M., Oikarinen E. Measuring House Price Bubbles," Real Estate Economics, American Real Estate and Urban Economics Association, 47(2). 2019: pp. 534-563.
26. Byeonghwa Park, Jae Kwon Bae. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. Expert Systems With Applications 42(6).2015: pp. 2928-2934.

27. Case, K. E., Shiller R. J. Prices of Single Family Homes Since 1970: New Indexes for Four Cities, *New England Economic Review* (1987), pp. 45–56.
28. Chica-Olmo J., Cano-Guervos R., Chica-Rivas M. Estimation of Housing Price Variations Using Spatio-Temporal Data. *Sustainability* 11(6), 2019: pp. 1-21.
29. Chi-man Hui E., Ziyong Wang. 2014. Market sentiment in private housing market. *Habitat International*, 44: pp. 375–385.
30. Clark S.P., Coggin, T. D. Was there a U.S. house price bubble? An econometric analysis using national and regional panel data. *The Quarterly Review of Economics and Finance*, Elsevier, 51(2). 2011: pp. 189-200.
31. Del Giudice V., De Paola P., Manganelli B., Forte F. The Monetary Valuation of Environmental Externalities through the Analysis of Real Estate Prices. *Sustainability* 2017, 9, 229.
32. Englund P., Quigley J.M., Redfearn C.L. The Choice of Methodology for Computing Housing Price Indexes: Comparisons of Temporal Aggregation and Sample Definition. *The Journal of Real Estate Finance and Economics* 19, pp. 91–112(1999)
33. Epley D. Assumptions and restrictions on the use of repeat sales to estimate residential price appreciation. *Journal of Real Estate Literature*. 2016; 24; 2: pp. 275–286.
34. Fabozzi, F. J., Xiao, K. The Timeline Estimation of Bubbles: The Case of Real Estate. *Real Estate Economics* 47(2). 2019: pp. 564-594.
35. Fernandez-Kranz, D., Hon, M.T. A cross-section analysis of the income elasticity of housing demand in Spain: Is there a real estate bubble? *Journal of real estate finance and economics*. 32(4). 2006: pp. 449-470.
36. Haizhen Wen, Yan Qin Zhang, Ling Zhang. Assessing amenity effects of urban landscapes on housing price in Hangzhou. *Urban Forestry & Urban Greening* 14(4), 2015: pp. 1017-1026.
37. Hakan Kusan, Osman Aytekin, Ilker Özdemir: The use of fuzzy logic in predicting house selling price. *Expert Syst. Appl.* 37(3). 2010: pp.1808-1813.
38. He Chengjie, Zhen Wang, Huaicheng Guo, Hu Sheng, Rui Zhou, Yonghui Yang. Driving Forces Analysis for Residential Housing Price in Beijing. *Procedia Environmental Sciences*, 2010: pp. 925 – 936.
39. Herath, S. K. & Maier, G. (2010). The hedonic price method in real estate and housing market research. A review of the literature. *Institute for Regional Development and Environment* (pp. 1-21). Vienna, Austria: University of Economics and Business
40. I-Cheng Yeh, Tzu-Kuang Hsu. Building real estate valuation models with comparative approach through case-based reasoning. *Appl. Soft Comput.* 65: pp. 260-271 (2018)
41. Jim, C. Y., & Chen, W. Y. (2006). Impacts of urban environmental elements on residential

- housing prices in Guangzhou (China). *Landscape and Urban Planning*, 78(4), pp. 422-434.
42. Jirong Gu, Mingcang Zhu, Liuguangyan Jiang. Housing price forecasting based on genetic algorithm and support vector machine. *Expert Syst. Appl.* 38(4): pp. 3383-3386 (2011).
 43. Kontrimas V., Verikas A.A. The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing* 11(1). 2011: pp. 443-448.
 44. Kouwenberg R., Zwinkels R. Forecasting the US housing market. *International Journal of Forecasting*, 30(3). 2014: pp. 415-425.
 45. Lughofer E., Trawinski B., Trawinski K., Kempa O., Lasota T. On employing fuzzy modeling algorithms for the valuation of residential premises. *Information Science* 181(23). 2011: pp. 5123-5142.
 46. Martin J.B., Muth R.F., Nourse H.O. A Regression Method for Real Estate. *Price Index Construction, Journal of the American Statistical Association*, 58:304 (1963), pp. 933-942.
 47. Peterson S., Flanagan A.B. Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research, American Real Estate Society*, 31(2). 2009: pp. 147-164.
 48. Phillips, P.C.B., Shi, S.-P. & Yu, J. Testing for Multiple Bubbles: Historical Episodes of Exuberance. *International Economic Review* 56(4). 2015: pp. 1043– 1078.
 49. Phillips, P.C.B., Shi, S.-P., & Yu, J. Testing for Multiple Bubbles: Limit Theory of Real Time Detectors. *International Economic Review* 56(4). 2015: pp. 1079– 1134.
 50. Plakandaras V., Gupta R., Gogas P., Papadimitriou T. Forecasting the U.S. Real House Price Index. *Economic Modelling, Elsevier*, 45(C). 2015: pp. 259-267.
 51. Poursaeed O., Matera T., Belongie S. Vision-based real estate price estimation. *Machine Vision and Applications* 29. 2018: pp. 667–676.
 52. Wilcox J.A. The Home Purchase Sentiment Index: A New Housing Indicator. *Business Economics* 50, pp. 178–190(2015).
 53. Xibin Wang, Junhao Wen, Yihao Zhang, Yubiao Wang. Real estate price forecasting based on SVM optimized by PSO. *Optik*, 125(3). 2014: pp. 1439-1443.
 54. Zhang, Yanbing & Hua, Xiuping & Zhao, Liang. Exploring determinants of housing prices: A case study of Chinese experience in 1999–2010. *Economic Modelling, Elsevier*, vol. 29(6). 2012: pp. 2349-2361.

Статистические сборники и отчеты

55. Отчет об определении кадастровой стоимости объектов недвижимости (за исключением земельных участков), расположенных на территории Санкт-Петербурга. КУГИ Правительства Санкт-Петербурга, том 2, раздел 2.3, СПб, 2012

http://rosreestr.ru/wps/portal/p/cc_ib_portal_services/cc_ib_ais_fdg/co/ Субъект РФ: С-Петербург. Виды объектов: Недвижимость, помещения. Отчет №32-1-0733/2012(2), Санкт-Петербург, 30.11.2012

56. Отчет об определении кадастровой стоимости объектов недвижимости на территории Санкт-Петербурга. КИО Правительства Санкт-Петербурга, том 1-5, СПб, 24.09.2018, утвержден приказом №130 КИО Правительства Санкт-Петербурга от 28.11.2018 г. <http://guko.commim.gov.spb.ru/>

Интернет-ресурсы и электронные базы данных

57. URL: <https://www.bn.ru/> (Дата обращения – 12.11.2019) - Сайт Бюллетеня недвижимости
58. URL: <https://www.oracle.com/ru/big-data/guide/what-is-big-data.html#close> (Дата обращения 04.04.2020) – официальный сайт Oracle в России и СНГ